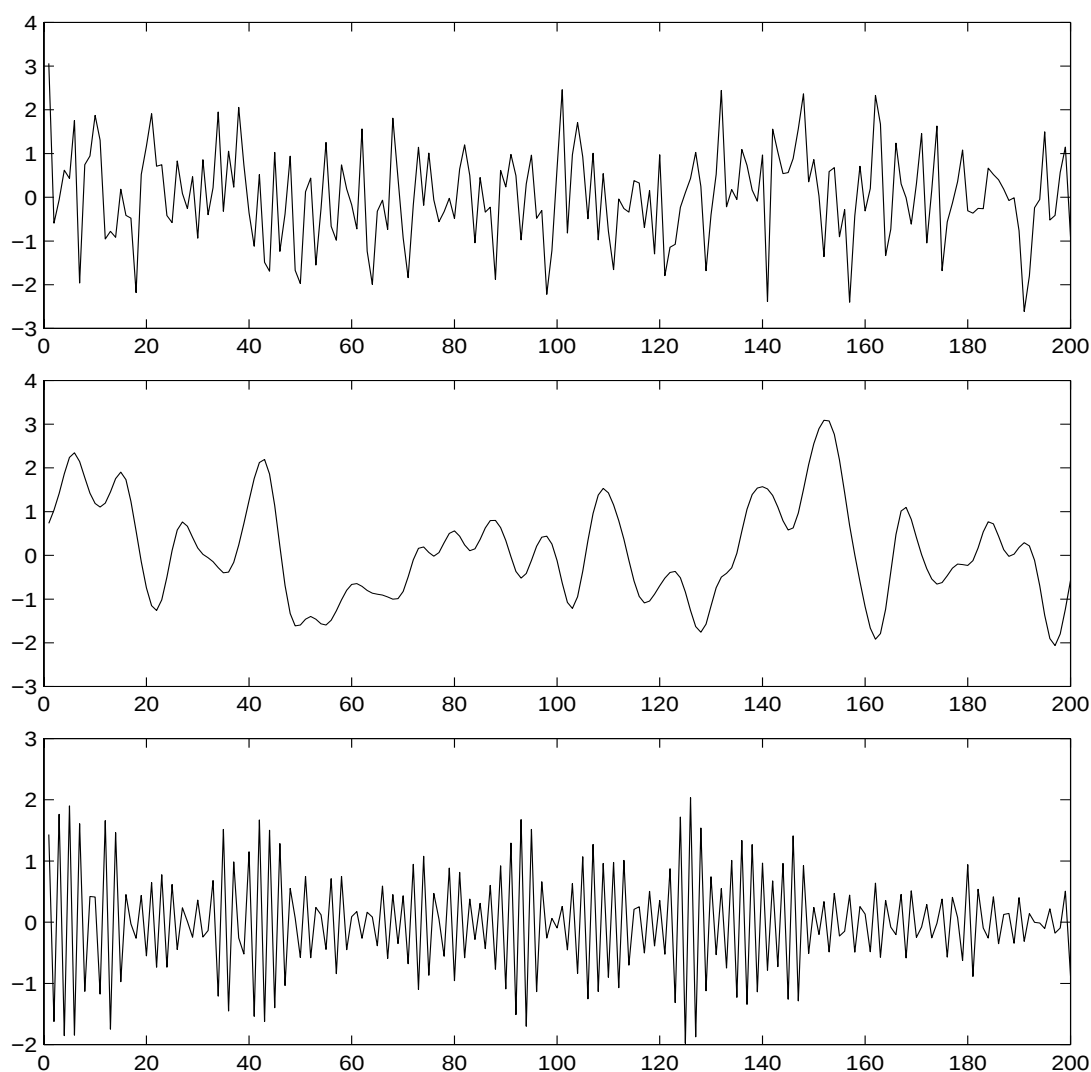


Mihai Ciuc

Constantin Vertan

PRELUCRAREA STATISTICĂ A SEMNALELOR



Editura MatrixRom
2005

Cuvânt înainte

Această lucrare reprezintă baza cursului de *Teoria transmisiunii informației 2*, curs de tradiție al Facultății de Electronică, Telecomunicații și Tehnologia Informației din Universitatea “Politehnica” București, obligatoriu pentru toți studenții anului III, indiferent de specializarea pe care o urmează. Structura prezentării urmărește, deci, fidel structura cursului, ale cărui baze au fost puse de domnul profesor Alexandru Spătaru și sunt expuse în lucrarea [2].

În plus, se presupune că cititorii sunt deja familiarizați cu o serie de noțiuni (precum algebră liniară, integrale multidimensionale, analiză Fourier, teoria sistemelor liniare, teoria informației) care sunt dobândite la cursurile urmate în primii doi ani de studiu (Algebră liniară, Analiză, Matematici speciale, Semnale, circuite și sisteme, Teoria Transmisiunii Informației 1). De asemenea, se presupune că cititorii sunt familiarizați cu noțiunea de probabilitate, (noțiune esențială pentru această lucrare) despre care sunt doar reamintite pe scurt elementele de bază.

Dorim să mulțumim D-lui Vasile Buzuloiu, profesor la Catedra de Electronică Aplicată și Ingineria Informației a Facultății de Electronică, Telecomunicații și Tehnologia Informației. Prezența continuă a Domniei Sale alături de noi de-a lungul ultimilor zece ani a contribuit esențial la formarea noastră (științifică și umană) astfel încât contribuția Domniei Sale la această lucrare este cu mult mai importantă decât ajutorul direct dat la redactarea și revizuirea ei.

De asemenea, dorim să mulțumim D-rei Lavinia Dârlea și D-lui Bogdan Ionescu, doctoranzi la Laboratorul de Analiza și Prelucrarea Imaginilor, care au avut bunăvoința și răbdarea de a parcurge minuțios materialul, ajutând, astfel, substanțial la îmbunătățirea formei lui.

Aprilie 2005,

Autorii

Cuprins

1	Introducere	1
2	Noțiuni de teoria probabilității	3
2.1	Introducere	3
2.2	Axiomele probabilității	4
2.3	Câmp de evenimente	6
2.4	Probabilități condiționate. Evenimente independente	7
3	Variabile aleatoare	9
3.1	Introducere	9
3.2	Evenimente generate de variabile aleatoare	10
3.3	Funcția de repartiție	11
3.4	Densitatea de probabilitate	12
3.5	Distribuții condiționate	15
3.6	Momente	16
3.7	Tipuri de distribuții	18
3.7.1	Distribuția uniformă	19
3.7.2	Distribuția gaussiană (normală)	20
3.7.3	Distribuția Rayleigh	20
3.8	Funcții de o variabilă aleatoare	20
3.9	Teorema de medie	29
4	Perechi de variabile aleatoare	31
4.1	Introducere	31
4.2	Funcția de repartiție de ordinul doi	31
4.3	Densitatea de probabilitate de ordinul doi	33
4.4	Distribuții condiționate	36
4.5	Variabile aleatoare independente	37
4.6	O funcție de două variabile aleatoare. Teorema limită centrală.	37
4.7	Teorema de medie	41
4.8	Momente	41
4.9	Dreapta de regresie	43
4.9.1	Coeficientul de corelație	46

4.10	Distribuția gaussiană de ordinul doi	47
5	Semnale aleatoare	49
5.1	Introducere	49
5.2	Caracterizarea statistică a semnalelor aleatoare	49
5.2.1	Caracterizarea de ordinul unu a semnalelor aleatoare	50
5.2.2	Caracterizarea de ordinul doi a semnalelor aleatoare	50
5.3	Semnale staționare	51
5.3.1	Staționaritate în sens strict	51
5.3.2	Staționaritate în sens larg	53
5.4	Proprietățile funcției de autocorelație	53
5.5	Funcția de intercorelație	55
5.6	Semnale ergodice	55
5.6.1	Medii temporale ale semnalului	55
5.6.2	Ergodicitate în sensul mediei și în sensul autocorelației	56
5.6.3	Teorema ergodicității mediei	56
5.7	Densitatea spectrală de putere	60
5.8	Teorema Wiener–Hincin	63
5.9	Densitatea spectrală de putere de interacțiune	66
5.10	Zgomotul alb	66
5.11	Trecerea semnalelor aleatoare prin filtre liniare	67
5.11.1	Sisteme liniare invariante în timp	68
5.11.2	Relații între mărimile statistice la trecerea prin filtre liniare	70
5.11.3	Trecerea semnalelor aleatoare prin FTJ ideal	71
5.11.4	Trecerea semnalelor aleatoare prin FTB ideal	74
5.12	Filtrul adaptat la semnal	78
5.12.1	Maximizarea RSZ prin filtrare liniară	80
6	Detecția semnalelor	83
6.1	Introducere	83
6.2	Criteriul de decizie Bayes: cazul observațiilor discrete	85
6.2.1	Statistica suficientă	90
6.2.2	Criteriul Bayes pentru zgomot alb gaussian	90
6.3	Criteriul Bayes: cazul observațiilor continue	93
6.3.1	Cazul zgomotului alb gaussian	95
7	Estimarea parametrilor	99
7.1	Introducere	99
7.2	Estimarea în sensul costului mediu minim	100
7.2.1	Estimarea în sensul funcției de cost pătratul erorii	101
7.2.2	Estimarea în sensul funcției de cost uniforme	102
7.2.3	Estimarea unui parametru gaussian în zgomot alb, gaussian	104

7.2.4	Estimarea în absența unui model statistic a priori	109
7.3	Evaluarea calității unui estimator	109
8	Semnale aleatoare în timp discret	111
8.1	Introducere	111
8.2	Matricea de autocorelație	112
8.3	Modele stochastice	114
8.3.1	Modelul AR	115
8.3.2	Modelul MA	117
8.3.3	Modelul ARMA	117
8.3.4	Ecuațiile Yule–Walker	118
9	Filtrarea optimală a semnalelor	123
9.1	Introducere	123
9.2	Principiul ortogonalității	124
9.3	Ecuațiile Wiener–Hopf	125
9.3.1	Ecuațiile Wiener–Hopf pentru filtru de tip FIR	126
9.4	Aplicații ale filtrării optimale a semnalelor	127
9.4.1	Atenuarea zgomotului	128
9.4.2	Predicția	129
10	Transformate unitare	133
10.1	Introducere	133
10.2	Transformata Karhunen–Loève	136
10.3	Aproximări practice ale transformatei KL	139
10.3.1	Transformata cosinus discretă	140
10.3.2	Transformata sinus discretă	143
11	Cuantizarea semnalelor	145
11.1	Introducere	145
11.2	Cuantizarea uniformă	147
11.3	Cuantizarea optimală Lloyd–Max	148
11.4	Compandarea	150
12	Aplicații în prelucrarea și analiza imaginilor	153
12.1	Introducere	153
12.2	Îmbunătățirea imaginilor	156
12.2.1	Egalizarea de histogramă	156
12.3	Segmentarea imaginilor	157
12.3.1	Segmentarea imaginilor ca problemă de cuantizare	160
12.3.2	Segmentarea imaginilor ca problemă de decizie	162
12.4	Compresia imaginilor cu transformate	164

A Impulsul Dirac	169
Bibliografie	171

Capitolul 1

Introducere

Spre deosebire de abordarea deterministă a prelucrării semnalelor, în care fiecare semnal este considerat o entitate de sine stătătoare, în prelucrarea statistică se ia în considerare faptul că un semnal purtător de informație aparține unei anumite *clase de semnale* (spre exemplu, semnal vocal etc.) clasă ce poate fi caracterizată prin anumiți parametri statistici. Scopul teoriei statistice a semnalelor este definirea acestor parametri și utilizarea lor pentru o prelucrare eficientă a semnalelor.

Prima parte a acestei lucrări este dedicată definirii mărimilor care caracterizează statistic semnalele aleatoare. În capitolul 2 sunt reluate pe scurt elementele de bază ale teoriei probabilității, pe care le vom utiliza în capitolul 3 pentru introducerea noțiunii de variabilă aleatoare: sunt definite mărimile statistice ce caracterizează variabilele aleatoare (funcție de repartiție, densitate de probabilitate, momente) și este discutată în detaliu problema modificării acestora prin aplicarea unei funcții cunoscute variabilei aleatoare. Capitolul 4 tratează problema caracterizării statistice comune a unei perechi de variabile aleatoare, cu ajutorul căreia se pot trage concluzii referitoare la gradul și tipul de dependență statistică între acestea.

Toate mărimile definite și toate rezultatele obținute pentru variabilele aleatoare vor servi în capitolul 5 pentru caracterizarea statistică a semnalelor aleatoare. Sunt definite noțiunile, extrem de importante, de staționaritate și ergodicitate a semnalelor. De asemenea, se definesc mărimi statistice care să caracterizeze semnalele din punct de vedere spectral și este demonstrată o teoremă fundamentală în prelucrarea semnalelor (teorema Wiener–Hincin) care face legătură între mărimile statistice temporale și spectrale. De asemenea, este studiată modificarea mărimilor statistice la trecerea semnalelor prin sisteme liniare, invariante în timp.

Partea a doua a lucrării prezintă problema prelucrării semnalelor aleatoare folosind caracteristicile statistice definite în prima parte. Capitolele 6 și 7 tratează probleme de decizie statistică: detecția semnalelor și estimarea parametrilor. În primul caz, este vorba despre identificarea cât mai precisă, în prezența zgomotului, a unui semnal dintr-o mulțime finită de semnale cunoscute (cu aplicație în extragerea informației în transmisiuni digitale), în timp ce în al doilea caz, problema este de a estima cât mai precis (de asemenea, în prezența zgomotului) a unui parametru necunoscut al unui semnal cunoscut. În

capitolul 8 se introduce reprezentarea în timp discret a semnalelor aleatoare și se discută noțiunea de model aleator pentru generarea semnalelor de interes. Capitolul 9 prezintă problema estimării optimale a formei semnalelor prin filtrare liniară, și sunt prezentate două aplicații de interes practic ridicat: reducerea de zgomot și predicția. În capitolul 10 este tratată problema reprezentării semnalelor în alte coordonate, cu aplicații imediate pentru analiza și compresia datelor. În sfârșit, în capitolul 11 este discutată problema cuantizării semnalelor, adică a transformării naturii valorilor semnalelor din continuu în discret.

A treia parte a lucrării (capitolul 12) propune o incursiune interesantă în lumea unei clase particulare de semnale aleatoare, și anume imaginile. Mai precis, sunt prezentate câteva aplicații ale prelucrării statistice a semnalelor în domeniul, extrem de actual, al prelucrării și analizei imaginilor digitale. Scopul principal al acestei părți este de a-i da cititorului, pe exemple concrete, o noțiune despre utilitatea tehnicilor descrise în această lucrare, noțiune care, datorită aridității prezentării (ce nu poate fi ocolită pe alocuri) riscă să treacă altminteri neobservată...

Capitolul 2

Noțiuni de teoria probabilității

2.1 Introducere

Probabilitățile sunt instrumentul matematic de bază pentru prezenta lucrare, drept pentru care vom relua în acest capitol elementele de bază din teoria probabilității ce ne vor fi de folos în continuare. Trebuie să precizăm că nu avem intenția de a face o prezentare exhaustivă a teoriei probabilităților, pentru aceasta cititorul fiind rugat să apeleze la una dintre numeroasele lucrări de referință din domeniu (spre exemplu [3]).

Asupra probabilității unui eveniment există o accepțiune foarte evidentă venită dinspre practică. Dacă repetăm un experiment de N ori, iar evenimentul A se produce de N_A ori, atunci probabilitatea $P(A)$ de producere a evenimentului A este:

$$P(A) = \lim_{N \rightarrow \infty} \frac{N_A}{N}. \quad (2.1)$$

Din păcate, această descriere intuitivă a probabilității unui eveniment, deși adevărată, este insuficient de precisă pentru a putea permite o construcție matematică pe seama ei, datele fiind astfel imposibil de manipulat. Această construcție matematică a teoriei probabilităților, de care avem nevoie în continuare, se face pornind de la trei proprietăți pe care le acceptăm fără demonstrație (axiome). Trebuie să remarcăm că deși abordarea e diferită, totuși există o potrivire perfectă a realității practice, descrise de relația (2.1), pe întregul eșafodaj matematic construit pe baza celor trei axiome.

Să începem prin a stabili terminologia folosită în teoria probabilităților. Fie $\Omega = \{\omega_1, \omega_2, \dots\}$ mulțimea tuturor rezultatelor posibile ale unui experiment. În teoria probabilității, elementele ω_i se numesc realizări experimentale, submulțimile $A \subset \Omega$ se numesc *evenimente*, mulțimea Ω se numește *evenimentul sigur*, mulțimea vidă $\emptyset$ se numește *evenimentul imposibil*, iar submulțimile formate din câte un singur element $\{\omega_i\}$ se numesc *evenimente elementare*. Două evenimente A și B se numesc *incompatibile* dacă $A \cap B = \emptyset$.

Dacă la o desfășurare a experimentului s-a observat ca rezultat al acestuia elementul ω_i , atunci se spune că evenimentul A s-a produs (sau a avut loc) dacă $\omega_i \in A$. De asemenea, dacă $\omega_i \notin B$, se spune că evenimentul B nu s-a produs. În acest sens, este evident că evenimentul Ω se produce la fiecare desfășurare a experimentului, iar evenimentul $\emptyset$ nu

se produce niciodată. De asemenea, două evenimente incompatibile nu se pot produce simultan niciodată.

Din definițiile date mai sus, rezultă că la o desfășurare a experimentului se produc concomitent mai multe evenimente. Într-adevăr, dacă se observă ω_i , se produc *toate* evenimentele $A \ni \omega_i$. Să exemplificăm acest lucru pe experimentul care constă în aruncarea zarului. În acest caz, Ω este mulțimea tuturor celor șase fețe ale zarului $\Omega = \{f_1, f_2, \dots, f_6\}$. Dacă zarul cade, la o aruncare, pe fața f_2 , atunci se produc în același timp evenimentele distincte $A = \{\text{față pară}\} = \{f_2, f_4, f_6\}$, $B = \{\text{față cu număr} \leq 3\} = \{f_1, f_2, f_3\}$, $C = \{f_2, f_3\}$, în timp ce evenimentele $D = \{\text{față impară}\} = \{f_1, f_3, f_5\}$, $E = \{\text{față cu număr} \geq 4\} = \{f_4, f_5, f_6\}$ etc. nu se produc¹.

Un alt exemplu în același sens este următorul: considerăm o cutie ce conține mai multe forme geometrice, caracterizate de culori, forme și dimensiuni diferite. De exemplu: “sferă mică și roșie”, “cub mare și verde”, “tetraedru mare și albastru” etc. În acest caz $\Omega = \{\text{mulțimea tuturor pieselor din cutie}\}$. Dacă la extragerea la întâmplare a unei forme din cutie se scoate, să zicem, sfera mare și galbenă, atunci au loc concomitent evenimentele $A = \{\text{sferă}\} = \{\text{mulțimea tuturor formelor de formă sferică}\}$, $B = \{\text{galben}\} = \{\text{mulțimea tuturor formelor de culoare galbenă}\}$ și $C = \{\text{mare}\} = \{\text{mulțimea tuturor formelor de dimensiune mare}\}$, în timp ce evenimentele $D = \{\text{albastru}\}$ și $E = \{\text{tetraedru}\}$ nu se produc.

2.2 Axiomele probabilității

Probabilitatea P este o măsură care se asociază evenimentelor $A \subset \Omega$ și care satisface următoarele axiome:

1. Probabilitatea oricărui eveniment este un număr pozitiv:

$$P(A) \geq 0 \quad \forall A.$$

2. Probabilitatea evenimentului sigur este egală cu unitatea:

$$P(\Omega) = 1.$$

3. Dacă două evenimente sunt incompatibile, atunci probabilitatea reuniunii celor două evenimente este egală cu suma probabilităților evenimentelor:

$$A \cap B = \emptyset \quad \Rightarrow \quad P(A \cup B) = P(A) + P(B).$$

Observație. Prin extinderea axiomei 3, se poate calcula $P(A \cup B \cup C) = P(A) + P(B) + P(C)$ pentru trei evenimente incompatibile două câte două și tot așa pentru patru ș.a.m.d., cu alte cuvinte, axioma 3 poate fi extinsă pentru orice număr finit de evenimente A_i , $i = 1 \dots, N$. Axioma 3 nu se mai poate extinde, însă, când $N \rightarrow \infty$; deci ea trebuie “completată” pentru a îngloba și cazul unui număr numărabil de mulțimi. Introducem, deci, și axioma 3a.

¹În total, sunt $2^5 = 32$ evenimente distincte care se produc și tot $2^5 = 32$ evenimente care nu se produc simultan, la o singură aruncare a zarului.

- (a) Pentru orice mulțime numărabilă de evenimente incompatibile două câte două $\{A_1, A_2, A_3, \dots\}$, cu $A_i \cap A_j = \emptyset$, dacă $i \neq j$, avem:

$$P\left(\bigcup_i A_i\right) = \sum_i P(A_i).$$

Proprietățile probabilității

Pe baza celor trei axiome se mai pot demonstra următoarele proprietăți ale probabilității pe care le vom numerota în continuarea axiomelor.

4. $P(\emptyset) = 0$.

Demonstrație. Avem $A = A \cup \emptyset$, $\forall A$. Pe de altă parte $A \cap \emptyset = \emptyset$, $\forall A$. Aplicând axioma 3, obținem:

$$P(A) = P(A \cup \emptyset) \stackrel{\text{ax.3}}{=} P(A) + P(\emptyset),$$

de unde rezultă, evident, proprietatea enunțată.

5. $P(\mathcal{C}(A)) = 1 - P(A)$, cu $\mathcal{C}(A) = \{\omega_i \in \Omega \mid \omega_i \notin A\}$.

Demonstrație. Avem $A \cup \mathcal{C}(A) = \Omega$ și $A \cap \mathcal{C}(A) = \emptyset$, de unde rezultă:

$$\underbrace{P(\Omega)}_1 = P(A \cup \mathcal{C}(A)) \stackrel{\text{ax.3}}{=} P(A) + P(\mathcal{C}(A)).$$

6. $P(A) \leq 1$, $\forall A$.

Demonstrație. Conform proprietății 5, $P(A) = 1 - P(\mathcal{C}(A))$, iar din axioma 2, $P(\mathcal{C}(A)) \geq 0$.

7. Pentru două evenimente oarecare A și B , avem:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Demonstrație. Putem scrie:

$$A = A \cap \Omega = A \cap (B \cup \mathcal{C}(B)) = (A \cap B) \cup (A \cap \mathcal{C}(B)),$$

și întrucât $(A \cap B) \cap (A \cap \mathcal{C}(B)) = \emptyset$ rezultă, aplicând axioma 3, că:

$$P(A) = P(A \cap B) + P(A \cap \mathcal{C}(B)). \quad (2.2)$$

Pe de altă parte, avem: $A \cup B = (A \cap \mathcal{C}(B)) \cup B$ (vezi figura 2.1), cu $(A \cap \mathcal{C}(B)) \cap B = \emptyset$, de unde:

$$P(A \cup B) = P(A \cap \mathcal{C}(B)) + P(B). \quad (2.3)$$

Eliminându-l pe $P(A \cap \mathcal{C}(B))$ din ecuațiile (2.2) și (2.3), obținem relația căutată.

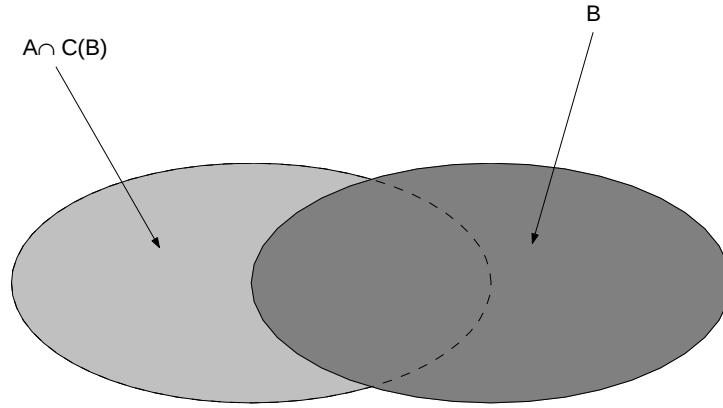


Figura 2.1: Ilustrarea grafică a relației $A \cup B = (A \cap \mathcal{C}(B)) \cup B$.

2.3 Câmp de evenimente

În general, nu suntem interesați (în unele cazuri nici măcar nu este posibil) să dăm probabilități tuturor evenimentelor ce se pot forma pe Ω . Desigur, alocarea unei probabilități pentru fiecare eveniment nu este o problemă atunci când Ω , de exemplu, e o mulțime cu șase elemente, caz în care $\mathcal{P}(\Omega)$ are $2^6 = 64$ de elemente distincte². Problema însă se complică pentru Ω având componente din ce în ce mai numeroase (Ω poate fi, evident, o mulțime cu o infinitate de elemente).

Dorim, deci, să alocăm probabilități unei mulțimi reduse de evenimente. Dorim, însă, ca mulțimea respectivă de submulțimi ale lui Ω să fie închisă la operațiunile elementare cu mulțimi, cum ar fi reuniunea, intersecția și complementarea. Această necesitate vine din faptul că dacă știm $P(A)$ și $P(B)$, atunci în general ne interesează să putem calcula și $P(A \cup B)$, și $P(\mathcal{C}(A))$, $P(A \cap B)$ etc, cu alte cuvinte, dacă A și B sunt în mulțimea restrânsă de submulțimi ale lui Ω cărora li se alocă probabilități, atunci $A \cup B$, $\mathcal{C}(A)$, $A \cap B$ etc. trebuie să fie tot în acea mulțime. Ajungem, astfel, la noțiunea de *câmp de evenimente*.

Definiție. Fie $\mathcal{K} \subset \mathcal{P}(\Omega)$ o mulțime de evenimente definite pe Ω . $\mathcal{K}$ se numește câmp de evenimente dacă:

1. $\forall A \in \mathcal{K} \Rightarrow \mathcal{C}(A) \in \mathcal{K}$;
2. $\forall A, B \in \mathcal{K} \Rightarrow A \cup B \in \mathcal{K}$.

Vom demonstra în continuare că un câmp de evenimente astfel definit are următoarele proprietăți:

3. $\forall A, B \in \mathcal{K} \Rightarrow A \cap B \in \mathcal{K}$.

²Pentru o mulțime oarecare A , mulțimea părților lui A , care se notează cu $\mathcal{P}(A)$ reprezintă mulțimea tuturor submulțimilor ce se pot forma cu elemente din A .

Demonstrație. Avem: $A, B \in \mathcal{K} \Rightarrow \mathcal{C}(A), \mathcal{C}(B) \in \mathcal{K} \Rightarrow \mathcal{C}(A) \cup \mathcal{C}(B) \in \mathcal{K} \Rightarrow \mathcal{C}(\mathcal{C}(A) \cup \mathcal{C}(B)) = A \cap B \in \mathcal{K}$

4. $\Omega \in \mathcal{K}$.

Demonstrație. Avem: $\forall A \in \mathcal{K} \Rightarrow \mathcal{C}(A) \in \mathcal{K} \Rightarrow A \cup \mathcal{C}(A) = \Omega \in \mathcal{K}$.

5. $\emptyset \in \mathcal{K}$.

Demonstrație. Avem: $\forall A \in \mathcal{K} \Rightarrow \mathcal{C}(A) \in \mathcal{K} \Rightarrow A \cap \mathcal{C}(A) = \emptyset \in \mathcal{K}$.

Definiție. Mulțimea $\mathcal{K} \subset \mathcal{P}(\Omega)$ se numește *câmp Borel* de evenimente dacă este un câmp de evenimente pentru care definiția 2 este valabilă pentru o infinitate numărabilă de mulțimi, deci se înlocuiește cu:

2'. $\forall A_i \in \mathcal{K}$ cu $i \in \mathbb{N}$, avem $\bigcup_i A_i \in \mathcal{K}$.

Pentru fixarea noțiunii de câmp de evenimente, să considerăm un exemplu practic. Să presupunem că dorim să construim câmpul de evenimente minimal $\mathcal{K}_1$ pe $\Omega = \{f_1, \dots, f_6\}$, care să conțină evenimentul $A = \{\text{față pară}\} = \{f_2, f_4, f_6\}$. Conform proprietăților 4 și 5 ale unui câmp de evenimente, $\mathcal{K}_1$ trebuie să conțină în afară de A și pe $\emptyset$ și Ω . De asemenea, el trebuie să-l conțină și pe $\mathcal{C}(A) = \{\text{față impară}\} = \{f_1, f_3, f_5\}$. Se observă că $\mathcal{K}_1 = \{\emptyset, \Omega, \{f_2, f_4, f_6\}, \{f_1, f_3, f_5\}\}$ reprezintă un câmp de evenimente, întrucât orice operație elementară cu mulțimi cu una sau două elemente din $\mathcal{K}_1$ conduce tot la un element din $\mathcal{K}_1$.

Câmpul minimal $\mathcal{K}_2$ care conține evenimentele elementare $B = \{f_1\}$ și $C = \{f_2\}$ este $\mathcal{K}_2 = \{\emptyset, \Omega, \underbrace{\{f_1\}}_B, \underbrace{\{f_2\}}_C, \underbrace{\{f_1, f_2\}}_{B \cup C}, \underbrace{\{f_1, f_3, f_4, f_5, f_6\}}_{\mathcal{C}(C)}, \underbrace{\{f_2, f_3, f_4, f_5, f_6\}}_{\mathcal{C}(B)}, \underbrace{\{f_3, f_4, f_5, f_6\}}_{\mathcal{C}(B \cup C)}\}$.

În concluzia acestei scurte introduceri în probabilități putem afirma că un experiment este complet determinat din punct de vedere probabilistic prin specificarea tripletului $(\Omega, \mathcal{K}, P)$, unde:

- Ω este mulțimea tuturor rezultatelor experimentale posibile;
- $\mathcal{K} \subset \mathcal{P}(\Omega)$ este câmpul de evenimente definit pe Ω ;
- P este mulțimea probabilităților asociate elementelor din $\mathcal{K}$.

2.4 Probabilități condiționate. Evenimente independente

Fie $A, B \subset \Omega$ două evenimente oarecare, cu $P(B) \neq 0$. Se definește probabilitatea evenimentului A condiționat de B ca fiind:

$$P(A|B) \triangleq \frac{P(A \cap B)}{P(B)}, \quad (2.4)$$

Probabilitatea de mai sus reprezintă probabilitatea de producere a evenimentului A considerând că evenimentul B s-a produs.

Exemplu. Considerăm experimentul ce constă în aruncarea zarului. Fie evenimentele $A = \{f_2\}$, respectiv $B = \{\text{față pară}\} = \{f_2, f_4, f_6\}$. Evident, avem $P(A) = \frac{1}{6}$, $P(B) = \frac{1}{2}$. Astfel, putem scrie:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}.$$

Rezultatul de mai sus era de așteptat, întrucât $P(A|B)$ reprezintă probabilitatea ca zarul să cadă pe fața f_2 în ipoteza producerii evenimentului B , adică știind faptul că zarul a căzut pe o față pară. Invers, putem scrie:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A)}{P(A)} = 1,$$

ceea ce iarăși e normal, întrucât $P(B|A)$ reprezintă probabilitatea ca zarul să cadă pe o față pară în ipoteza că el a căzut pe fața f_2 .

Definiție. Două evenimente A și B se numesc independente dacă:

$$P(A \cap B) = P(A)P(B). \quad (2.5)$$

Ținând cont de (2.4), rezultă că dacă A și B sunt independente, avem:

$$P(A|B) = \frac{P(A)P(B)}{P(B)} = P(A), \quad (2.6)$$

respectiv

$$P(B|A) = \frac{P(A)P(B)}{P(A)} = P(B), \quad (2.7)$$

ceea ce justifică definiția (2.5). Relațiile de mai sus se interpretează în sensul că dacă două evenimente sunt independente, atunci producerea oricăruia dintre ele nu influențează probabilitatea de producere a celuilalt.

Capitolul 3

Variabile aleatoare

3.1 Introducere

Prin definiție, o variabilă aleatoare este o funcție care asociază fiecărui element $\omega_i \in \Omega$ un număr real. Vom nota variabilele aleatoare cu litere grecești: $\xi, \eta, \zeta, \dots$. Putem, deci, scrie că:

$$\xi: \Omega \rightarrow \mathbb{R}. \quad (3.1)$$

De fiecare dată când se desfășoară experimentul și observăm ca rezultat al acestuia pe ω_k , valoarea asociată $\xi(\omega_k) \stackrel{\text{not}}{=} \xi^{(k)}$ se numește *realizarea particulară* a lui ξ .

Exemple.

1. pentru experimentul care constă în aruncarea zarului, putem defini pe $\Omega = \{f_1, \dots, f_6\}$ mai multe variabile aleatoare, ca de exemplu:

- $\xi(f_i) = i, \forall i \in \{1, \dots, 6\}$, variabila aleatoare care asociază fiecărei fețe a zarului numărul de pe ea;
- $\eta(f_i) = \begin{cases} 0 & \text{dacă } i = 2k \\ 1 & \text{dacă } i = 2k + 1 \end{cases}$, variabila aleatoare care ia valoarea 0 în cazul în care cade față pară, și 1 pentru față impară.

2. pentru experimentul care constă în alegerea la întâmplare a unui rezistor dintr-o cutie de N rezistoare r_i ieșite de pe aceeași linie de fabricație, atunci $\Omega = \{r_1, \dots, r_N\}$, pe care se pot defini, de asemenea, mai multe variabile de interes practic, cum ar fi:

- $\xi(r_i) = R(r_i)$, adică ξ este valoarea rezistenței electrice a rezistorului r_i ;
- $\eta(r_i) = C(r_i)$, adică η este valoarea capacității parazite a rezistorului r_i .

3.2 Evenimente generate de variabile aleatoare

În general, în practică este de interes cunoașterea probabilității ca o variabilă aleatoare să ia valori într-un anumit interval din $\mathbb{R}$, sau să fie mai mică decât o anumită valoare reală. Spre exemplu, în experimentul 2 descris mai sus, producătorul rezistoarelor respective este interesat să cunoască probabilitatea ca valoarea reală a rezistenței rezistorului (respectiv, variabila aleatoare ξ definită în exemplul 2) să ia valori în intervalul de toleranță specificat, și exemplele pot continua la nesfârșit.

După cum știm, însă, din capitolul precedent, putem alocă probabilități numai evenimentelor $A \subset \Omega$. Vom vedea în continuare cum putem să generăm, pornind de la aceste evenimente de probabilitate cunoscută, evenimente legate de variabilele aleatoare.

Fie ξ o variabilă aleatoare definită pe Ω și fie $x_1, x_2 \in \mathbb{R}$. Atunci, definim evenimentul $\{x_1 \leq \xi \leq x_2\}$ ca fiind:

$$\{x_1 \leq \xi \leq x_2\} = \left\{ \omega_i \in \Omega \mid x_1 \leq \xi(\omega_i) \leq x_2 \right\}. \quad (3.2)$$

Spre exemplu, pentru ξ definită la exemplul 1, putem scrie:

$$\begin{aligned} \{\xi \geq 2, 5\} &= \left\{ f_i \in \Omega \mid \xi(f_i) \geq 2, 5 \right\} = \{f_3, f_4, f_5, f_6\} \\ \{2 < \xi \leq 4\} &= \left\{ f_i \in \Omega \mid 2 < \xi(f_i) \leq 4 \right\} = \{f_3, f_4\} \\ \{\xi < 0, 5\} &= \left\{ f_i \in \Omega \mid \xi(f_i) < 0, 5 \right\} = \emptyset \\ \{-5 \leq \xi \leq 8, 3\} &= \left\{ f_i \in \Omega \mid -5 \leq \xi(f_i) \leq 8, 3 \right\} = \{f_1, \dots, f_6\} = \Omega \end{aligned}$$

Se observă că pentru orice $x_1, x_2 \in \mathbb{R}$, $\{x_1 < \xi \leq x_2\}$, $\{\xi < x_1\}$, $\{\xi \geq x_2\}$ etc. nu sunt altceva decât submulțimi ale lui Ω , ceea ce justifică definirea lor ca evenimente. Înseamnă că putem afla probabilitatea acestor evenimente, în măsura în care știm probabilitățile asociate evenimentelor definite pe Ω . Avem, deci, dreptul să vorbim de probabilitatea ca o variabilă aleatoare să ia valori într-un interval oarecare din $\mathbb{R}$, sau, prin extensie, într-o reuniune finită sau numărabilă de intervale din $\mathbb{R}$.

Este evident că pe $\mathbb{R}$ se pot defini submulțimi $A \subset \mathbb{R}$ care să nu fie intervale sau reuniuni de intervale, dar în majoritatea cazurilor nu este de interes practic determinarea probabilității $P(\xi \in A)$ pentru aceste submulțimi (și dăm numai un exemplu ilustrativ pentru aceasta: $A = \mathbb{R} \setminus \mathbb{Q}$), lăsând la o parte faptul că în unele cazuri respectivele probabilități nici nu pot fi calculate.

Ne mulțumim, deci, cu determinarea probabilităților $P(\xi \in A)$, cu $A \in \mathcal{I}$, unde $\mathcal{I} = \{\text{mulțimea tuturor intervalelor din } \mathbb{R} \text{ închise, deschise, sau semideschise și a tuturor reuniunilor unui număr finit sau numărabil de intervale}\}$. Este evident că $\mathcal{I}$ reprezintă un câmp de evenimente, fiind închis la complementare și reuniune.

Deși o astfel de descriere probabilistică a variabilelor aleatoare pare extrem de complicată, ea nu este așa. Vom arăta în paragraful următor că toate probabilitățile menționate

mai sus pot fi deduse pornind de la expresia unei singure funcții, numită *funcție de repartiție* a variabilei aleatoare.

3.3 Funcția de repartiție

Fie ξ o variabilă aleatoare. Se definește funcția ei de repartiție $F_\xi: \mathbb{R} \rightarrow [0, 1]$ cu expresia dată de:

$$F_\xi(x) \triangleq P(\xi \leq x) \quad (3.3)$$

Proprietățile funcției de repartiție

1. $F_\xi(-\infty) = P(\xi \leq -\infty) = 0$.
2. $F_\xi(\infty) = P(\xi \leq \infty) = 1$.
3. F_ξ e o funcție crescătoare: $\forall x_1, x_2 \in \mathbb{R}, x_1 < x_2 \Rightarrow F_\xi(x_1) \leq F_\xi(x_2)$.

Demonstrație. $F_\xi(x_2) = P(\xi \leq x_2) = P((\xi \leq x_1) \cup (x_1 < \xi \leq x_2))$. Având în vedere că evenimentele $\{\xi \leq x_1\}$ și $\{x_1 < \xi \leq x_2\}$ sunt incompatibile, putem scrie că:

$$F_\xi(x_2) = P(\xi \leq x_1) + P(x_1 < \xi \leq x_2) = F_\xi(x_1) + \underbrace{P(x_1 < \xi \leq x_2)}_{\geq 0},$$

ceea ce completează demonstrația.

4. $P(x_1 < \xi \leq x_2) = F_\xi(x_2) - F_\xi(x_1)$.

Această relație nu este decât relația finală din demonstrația proprietății 3 pusă sub o altă formă. O menționăm, însă, separat, pentru că este relația care justifică cele afirmate mai sus, cum că funcția de repartiție oferă suficientă informație pentru a calcula $P(\xi \in A)$, $\forall A \in \mathcal{I}$. Este evident că pentru orice interval sau reuniune de intervale poate fi scrisă o relație de genul celei de mai sus, care face apel numai la valorile funcției de repartiție în capetele intervalului(elor) respectiv(e).

Exemplu. Să calculăm spre exemplificare funcția de repartiție a variabilei aleatoare ξ definite în exemplul 1 de la începutul capitolului. Având în vedere natura discretă a lui ξ (care poate lua, conform definiției, numai valorile $1, 2, \dots, 6$), vom trata problema

calculului funcției de repartiție pe intervale. Astfel:

$$\begin{aligned}
 \forall x \in (-\infty, 1), F_{\xi}(x) &= P(\xi \leq x) = P(\emptyset) = 0 \\
 \forall x \in [1, 2), F_{\xi}(x) &= P(\xi \leq x) = P(\{f_1\}) = \frac{1}{6} \\
 \forall x \in [2, 3), F_{\xi}(x) &= P(\xi \leq x) = P(\{f_1, f_2\}) = \frac{2}{6} \\
 \forall x \in [3, 4), F_{\xi}(x) &= P(\xi \leq x) = P(\{f_1, f_2, f_3\}) = \frac{3}{6} \\
 \forall x \in [4, 5), F_{\xi}(x) &= P(\xi \leq x) = P(\{f_1, f_2, f_3, f_4\}) = \frac{4}{6} \\
 \forall x \in [5, 6), F_{\xi}(x) &= P(\xi \leq x) = P(\{f_1, f_2, f_3, f_4, f_5\}) = \frac{5}{6} \\
 \forall x \in [6, \infty), F_{\xi}(x) &= P(\xi \leq x) = P(\Omega) = 1
 \end{aligned}$$

Graficul funcției de repartiție este prezentat în figura 3.1

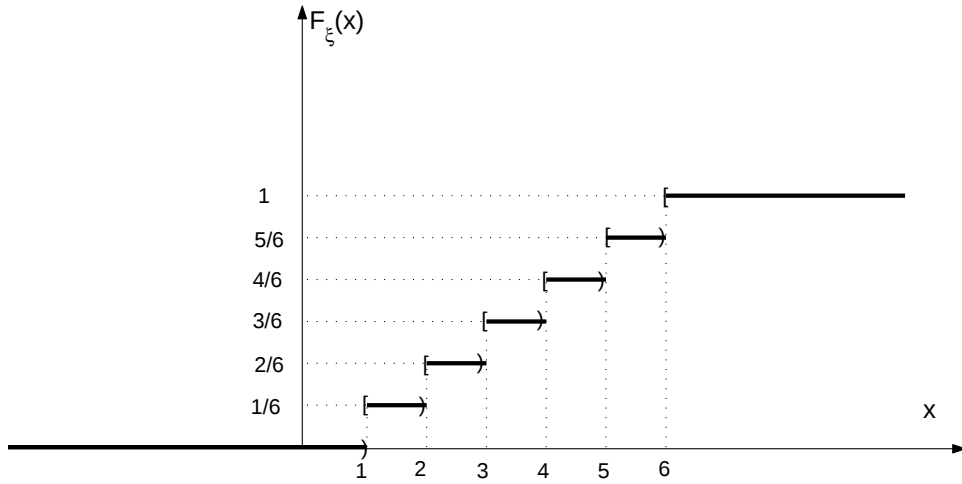


Figura 3.1: Funcția de repartiție a variabilei aleatoare “numărul de pe fața zarului”.

Definiție. O variabilă aleatoare se numește *continuuă* dacă funcția ei de repartiție este continuă pe $\mathbb{R}$, se numește *discretă* dacă funcția ei de repartiție este o funcție în trepte, de genul celei din figura 3.1, și se numește *mixtă* dacă funcția ei de repartiție prezintă discontinuități, fără a fi, totuși, o funcție în trepte.

3.4 Densitatea de probabilitate

Se definește densitatea de probabilitate w_{ξ} a unei variabile aleatoare ξ ca fiind derivata funcției sale de repartiție:

$$w_{\xi}(x) \triangleq \frac{dF_{\xi}(x)}{dx}. \quad (3.4)$$

Dacă explicităm derivata în relația de mai sus, obținem:

$$w_{\xi}(x) = \lim_{\Delta x \rightarrow 0} \frac{F_{\xi}(x + \Delta x) - F_{\xi}(x)}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{P(x < \xi \leq x + \Delta x)}{\Delta x}. \quad (3.5)$$

În calculul de mai sus am ținut cont de proprietatea 4 a funcției de repartiție. Se poate, deci, scrie:

$$w_{\xi}(x)\Delta x \underset{\Delta x \searrow 0}{\approx} P(x < \xi \leq x + \Delta x), \quad (3.6)$$

relație care justifică denumirea de densitate de probabilitate dată funcției w_{ξ} . Într-adevăr, se observă că valoarea funcției w_{ξ} într-un punct x permite calculul probabilității ca variabila aleatoare ξ să ia valori într-un interval infinitesimal $(x, x + \Delta x]$, această probabilitate fiind dată de aria de sub graficul lui w_{ξ} delimitată de intervalul $(x, x + \Delta x]$ (vezi figura 3.2).

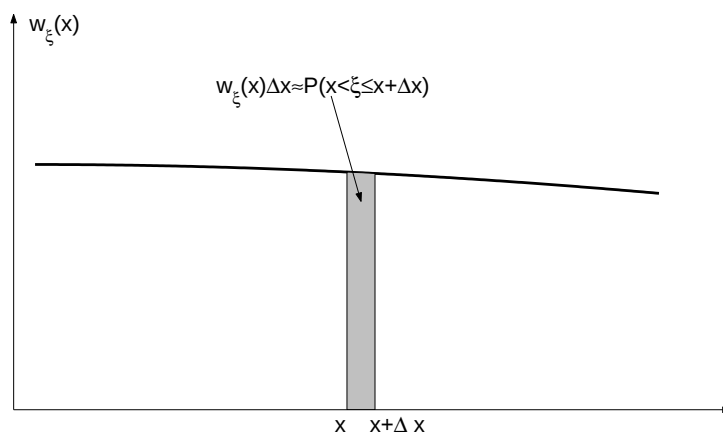


Figura 3.2: Probabilitatea este dată de aria de sub graficul densității de probabilitate.

Necesitatea utilizării funcției de *densitate* de probabilitate este dictată de faptul că, în cazul în care variabila aleatoare în discuție e continuă, avem $P(\xi = x) = 0 \ \forall x \in \mathbb{R}^1$.

Proprietățile densității de probabilitate

1. $w_{\xi}(x) \geq 0$

Demonstrație. w_{ξ} este derivata lui F_{ξ} care este funcție crescătoare.

2. $F_{\xi}(x) = \int_{-\infty}^x w_{\xi}(u) du.$

Demonstrație. Se ține cont de faptul că F_{ξ} este primitiva lui w_{ξ} , și că $F_{\xi}(-\infty) = 0$.

¹Pentru justificarea acestei afirmații, să ne imaginăm un zar cu $N = 20$ fețe simetrice (făcând abstracție de faptul că o astfel de figură geometrică nu există în realitate). În acest caz, probabilitatea de apariție a fiecărei fețe este $\frac{1}{20}$. Pe măsură ce N crește, scade probabilitatea fiecărei fețe. La limită, când $N \rightarrow \infty$, zarul se transformă într-o sferă, fața devine punct, iar probabilitatea fiecărei fețe este $\frac{1}{\infty}$, adică 0.

3. Condiția de normare: $\int_{-\infty}^{\infty} w_{\xi}(x)dx = 1$.

Demonstrație. Se scrie proprietatea 2 în punctul $x = \infty$.

4. Pentru $x_1, x_2 \in \mathbb{R}$, avem $\int_{x_1}^{x_2} w_{\xi}(x)dx = P(x_1 < \xi \leq x_2)$.

Demonstrație. Se utilizează proprietatea 2 a densității de probabilitate și proprietatea 4 a funcției de repartiție.

Această relație ne arată că aria de sub graficul densității de probabilitate delimitată de un interval reprezintă probabilitatea ca variabila aleatoare să ia valori în intervalul respectiv.

Exemplu. Să calculăm densitatea de probabilitate a variabilei ξ a cărei funcție de repartiție este ilustrată în figura 3.1.

După cum se observă, funcția de repartiție este constantă peste tot în $\mathbb{R}$, mai puțin în punctele $x = 1, \dots, 6$, unde are salturi bruște. În derivata funcției de repartiție, respectiv în densitatea de probabilitate w_{ξ} , în aceste puncte vor apărea impulsuri Dirac² de arii egale cu $\frac{1}{6}$, întrucât, în fiecare punct $i = 1, \dots, 6$, funcția F_{ξ} efectuează un salt de amplitudine $\frac{1}{6}$. Densitatea de probabilitate a lui ξ (prezentată grafic în figura 3.3) se poate scrie, deci:

$$w_{\xi}(x) = \sum_{i=1}^6 \frac{1}{6} \delta(x - i).$$

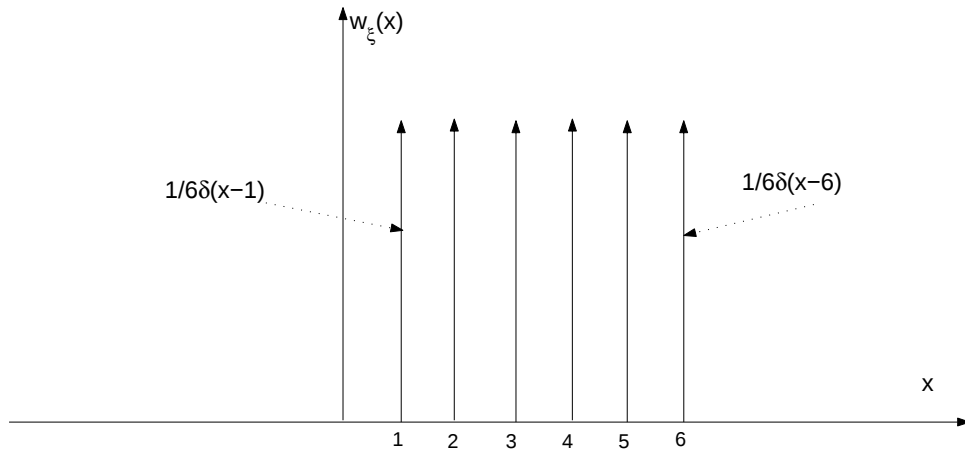


Figura 3.3: Densitatea de probabilitate a variabilei aleatoare “numărul de pe fața zarului”.

După cum este discutat în detaliu în anexa A, semnificația practică a unui impuls Dirac este aceea de arie nenulă “concentrată” într-un singur punct. Așadar, apariția

²Pentru mai multe detalii despre impulsul Dirac, a se consulta anexa A a acestei lucrări.

unui impuls Dirac de arie A într-un punct x_0 al unei funcții densitate de probabilitate w_ξ semnifică faptul că variabila aleatoare respectivă are probabilitate nenulă să ia exact valoarea x_0 , probabilitate care este egală cu aria impulsului respectiv³: $P(\xi = x_0) = A$.

3.5 Distribuții condiționate

Fie ξ o variabilă aleatoare și A un eveniment oarecare. Prin analogie cu (2.4), se definește funcția de repartiție a lui ξ condiționată de A :

$$F_{\xi|A}(x) = P((\xi \leq x)|A) = \frac{P((\xi \leq x) \cap A)}{P(A)}. \quad (3.7)$$

În mod similar cu (3.4), se definește densitatea de probabilitate condiționată:

$$w_{\xi|A}(x) = \frac{dF_{\xi|A}(x)}{dx} = \lim_{\Delta x \rightarrow 0} \frac{P((x < \xi \leq x + \Delta x)|A)}{\Delta x}. \quad (3.8)$$

Exemplu. Să considerăm o variabilă aleatoare ξ oarecare, și să considerăm evenimentul $A = \{a < \xi \leq b\}$, cu $a < b$ oarecare. Să calculăm într-o primă instanță funcția de repartiție condiționată $F_{\xi|A}(x)$. Avem, conform (3.7):

$$F_{\xi|\{a < \xi \leq b\}}(x) = \frac{P((\xi \leq x) \cap (a < \xi \leq b))}{P(a < \xi \leq b)} = \frac{P((\xi \leq x) \cap (a < \xi \leq b))}{F_\xi(b) - F_\xi(a)}. \quad (3.9)$$

- Pentru $x < a$, avem $\{\xi \leq x\} \cap \{a < \xi \leq b\} = \emptyset$, și deci:

$$F_{\xi|\{a < \xi \leq b\}}(x) = 0.$$

- Pentru $x \in (a, b]$ avem $\{\xi \leq x\} \cap \{a < \xi \leq b\} = \{a < \xi \leq x\}$, iar (3.9) se scrie:

$$F_{\xi|\{a < \xi \leq b\}}(x) = \frac{P(a < \xi \leq x)}{F_\xi(b) - F_\xi(a)} = \frac{F_\xi(x) - F_\xi(a)}{F_\xi(b) - F_\xi(a)}.$$

- Pentru $x > b$, avem $\{\xi \leq x\} \cap \{a < \xi \leq b\} = \{a < \xi \leq b\}$, de unde:

$$F_{\xi|\{a < \xi \leq b\}}(x) = \frac{P(a < \xi \leq b)}{F_\xi(b) - F_\xi(a)} = 1.$$

Cât despre densitatea de probabilitate condiționată, aplicând (3.8), avem:

$$w_{\xi|\{a < \xi \leq b\}}(x) = \begin{cases} 0 & \text{dacă } x < a \\ \frac{w_\xi(x)}{F_\xi(b) - F_\xi(a)} & \text{dacă } a < x \leq b \\ 0 & \text{dacă } x > b. \end{cases} \quad (3.10)$$

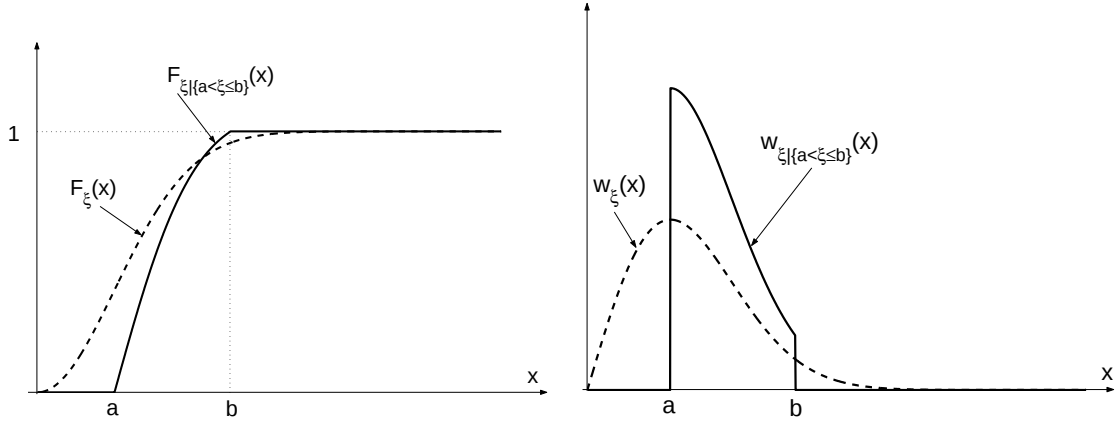


Figura 3.4: Distribuții condiționate.

În figura 3.4 sunt ilustrate funcția de repartiție și densitatea de probabilitate condiționate pentru o formă arbitrară a lui $F_\xi(x)$.

Să observăm din exemplul prezentat că atât funcția de repartiție condiționată cât și densitatea de probabilitate condiționată respectă toate proprietățile unei funcții de repartiție (descrise în paragraful 3.3) respectiv ale unei densități de probabilitate (paragraful 3.4). Această observație poate fi demonstrată pentru cazul general, pornind de la definițiile (3.7) și (3.8).

3.6 Momente

Fie variabila aleatoare ξ având distribuția w_ξ^4 . Se definește momentul necentrat de ordin k al variabilei aleatoare ξ și se notează cu $m_\xi^{(k)}$ sau cu $\bar{\xi}^k$ valoarea integralei:

$$m_\xi^{(k)} \stackrel{\text{not}}{=} \bar{\xi}^k = \int_{-\infty}^{\infty} x^k w_\xi(x) dx \quad k = 1, 2, \dots \quad (3.11)$$

Facem precizarea că pentru anumite variabile aleatoare, începând cu un anumit ordin k , momentele de ordin mai mare decât k pot fi infinite⁵.

³Aceasta justifică definiția variabilelor aleatoare discrete de la pagina 12: dacă funcția de repartiție e în trepte, atunci densitatea de probabilitate e o succesiune de impulsuri Dirac, deci este posibil numai un număr finit sau numărabil de valori pentru variabilele aleatoare respective.

⁴În restul lucrării, vom folosi termenul “distribuție” pentru a desemna o densitate de probabilitate.

⁵De exemplu, pentru variabila aleatoare cu distribuția.

$$w_\xi(x) = \begin{cases} \frac{1}{x^2} & \text{dacă } x \geq 1 \\ 0 & \text{în rest} \end{cases}$$

chiar și momentul de ordin unu este infinit.

De o mare importanță în studiul variabilelor aleatoare sunt momentele necentrate de ordin unu și doi. Astfel, momentul necentrat de ordin unu, dat de:

$$\bar{\xi} = \int_{-\infty}^{\infty} x w_{\xi}(x) dx \quad (3.12)$$

se numește *media* variabilei aleatoare.

Momentul necentrat de ordin doi:

$$\bar{\xi}^2 = \int_{-\infty}^{\infty} x^2 w_{\xi}(x) dx \quad (3.13)$$

se numește *media pătratică* a variabilei aleatoare.

Considerăm utilă în acest punct o discuție referitoare la formulele de definiție a mediei și mediei pătratice de mai sus. În general, există o accepțiune foarte clară asupra noțiunii de medie a unei variabile aleatoare, ca fiind media aritmetică a unui număr (de preferință cât mai mare) de realizări particulare ale variabilei aleatoare respective. După cum vom arăta în exemplul următor, această accepțiune nu contravine în nici un fel formulei (3.12). Revenind la experimentul 2 descris în paragraful 3.1 să presupunem că dorim să estimăm valoarea mediei a rezistenței electrice a rezistoarelor din cutia respectivă. În acest sens, măsurăm toate valorile rezistențelor, și calculăm valoarea medie aritmetică a acestora R_{med} . Să presupunem că, dispunând de un ohmmetru nu foarte precis, nu reținem decât valoarea întreagă a rezistențelor (deși, teoretic, valoarea respectivă este de natură continuă). Să presupunem că măsurăm următoarele valori pentru întreg setul de $N = 1825$ rezistoare:

Valoare rezistență R_i	19Ω	20Ω	21Ω	22Ω	23Ω	24Ω	25Ω
Număr rezistoare	89	211	432	501	342	171	79

Deci, valoarea medie a rezistenței R_{med} este:

$$\begin{aligned} R_{\text{med}} &= \frac{1}{N} \sum_{i=1}^N R(r_i) = \frac{89 \cdot 19\Omega + 211 \cdot 20\Omega + \dots + 79 \cdot 25\Omega}{1825} \\ &= 19\Omega \frac{89}{1825} + 20\Omega \frac{211}{1825} + \dots + 25\Omega \frac{79}{1825}. \end{aligned} \quad (3.14)$$

Rapoartele puse în evidență în relația de mai sus reprezintă probabilitățile diverselor valori posibile ale rezistenței electrice determinate *experimental*, pe setul de date disponibil (vezi relația (2.1)). Cu alte cuvinte $P(19\Omega) \approx \frac{89}{1825}$, $P(20\Omega) \approx \frac{211}{1825}$ etc. Putem, deci, rescrie pe (3.14) ca:

$$R_{\text{med}} \approx \sum_{R_i=19\Omega}^{25\Omega} R_i P(R_i). \quad (3.15)$$

Considerând valoarea rezistenței continuă (cum, de fapt, și este) și nu discretă (cum am fost forțați să o considerăm din lipsă de mijloace materiale suficient de sensibile), variabila discretă R_i se transformă în R , continuă, apoi suma $\sum_{R_i}$ devine o integrală $\int_R$, iar probabilitățile $P(R_i)$ se înlocuiesc cu $w(R)dR$.

Putem, deci, scrie pe (3.15) ca:

$$R_{\text{med}} \approx \int_{\mathbb{R}} R w(R) dR. \quad (3.16)$$

Relația (3.16) arată faptul că media aritmetică a unui număr de realizări particulare nu este altceva decât un *estimat* al valorii medii (3.12), pe care-l facem în absența cunoștințelor despre mărimile statistice (densitate de probabilitate, etc.) ale variabilei aleatoare respective!

Se definesc, de asemenea, momentele centrate de ordin k ale variabilei aleatoare, notate cu $M_\xi^{(k)}$, ca fiind:

$$M_\xi^{(k)} \stackrel{\text{not}}{=} \overline{(\xi - \bar{\xi})^k} = \int_{-\infty}^{\infty} (x - \bar{\xi})^k w_\xi(x) dx \quad k = 2, 3, \dots \quad (3.17)$$

Momentul centrat de ordinul doi:

$$M_\xi^{(2)} \stackrel{\text{not}}{=} \overline{(\xi - \bar{\xi})^2} = \int_{-\infty}^{\infty} (x - \bar{\xi})^2 w_\xi(x) dx \quad (3.18)$$

se numește *varianța* variabilei aleatoare, iar rădăcina pătrată a acestuia se notează cu σ_ξ și se numește *dispersia* variabilei aleatoare:

$$\sigma_\xi = \sqrt{M_\xi^{(2)}}. \quad (3.19)$$

Dispersia unei variabile aleatoare (numită de asemenea și *abatere medie pătratică*) măsoară gradul de împrăștiere al valorilor variabilei aleatoare față de valoarea medie. Cu cât dispersia este mai mică (sau mai mare), cu atât scade (respectiv crește) probabilitatea ca o realizare particulară a variabilei aleatoare să ia o valoare care să difere puternic de valoarea medie

În practică, se obișnuiește a se nota varianța ca fiind pătratul dispersiei:

$$M_\xi^{(2)} \stackrel{\text{not}}{=} \sigma_\xi^2. \quad (3.20)$$

Pornind de la formula varianței (3.18), se poate demonstra o relație simplă între aceasta pe de-o parte și media și media pătratică a variabilei aleatoare pe de altă parte:

$$\begin{aligned} \sigma_\xi^2 &= \int_{-\infty}^{\infty} (x^2 - 2x\bar{\xi} + \bar{\xi}^2) w_\xi(x) dx \\ &= \underbrace{\int_{-\infty}^{\infty} x^2 w_\xi(x) dx}_{\bar{\xi}^2} - 2\bar{\xi} \underbrace{\int_{-\infty}^{\infty} x w_\xi(x) dx}_{\bar{\xi}} + \bar{\xi}^2 \underbrace{\int_{-\infty}^{\infty} w_\xi(x) dx}_1 \\ &= \bar{\xi}^2 - 2\bar{\xi}^2 + \bar{\xi}^2 = \bar{\xi}^2 - \bar{\xi}^2. \end{aligned} \quad (3.21)$$

3.7 Tipuri de distribuții

În acest paragraf, vom prezenta pe scurt diverse tipuri de distribuții des întâlnite în prelucrarea semnalelor:

3.7.1 Distribuția uniformă

O variabilă aleatoare ξ are o distribuție uniformă în intervalul $[a, b]$ dacă densitatea ei de probabilitate este de forma:

$$w_{\xi}(x) = \begin{cases} \frac{1}{b-a} & \text{dacă } a \leq x \leq b \\ 0 & \text{în rest.} \end{cases} \quad (3.22)$$

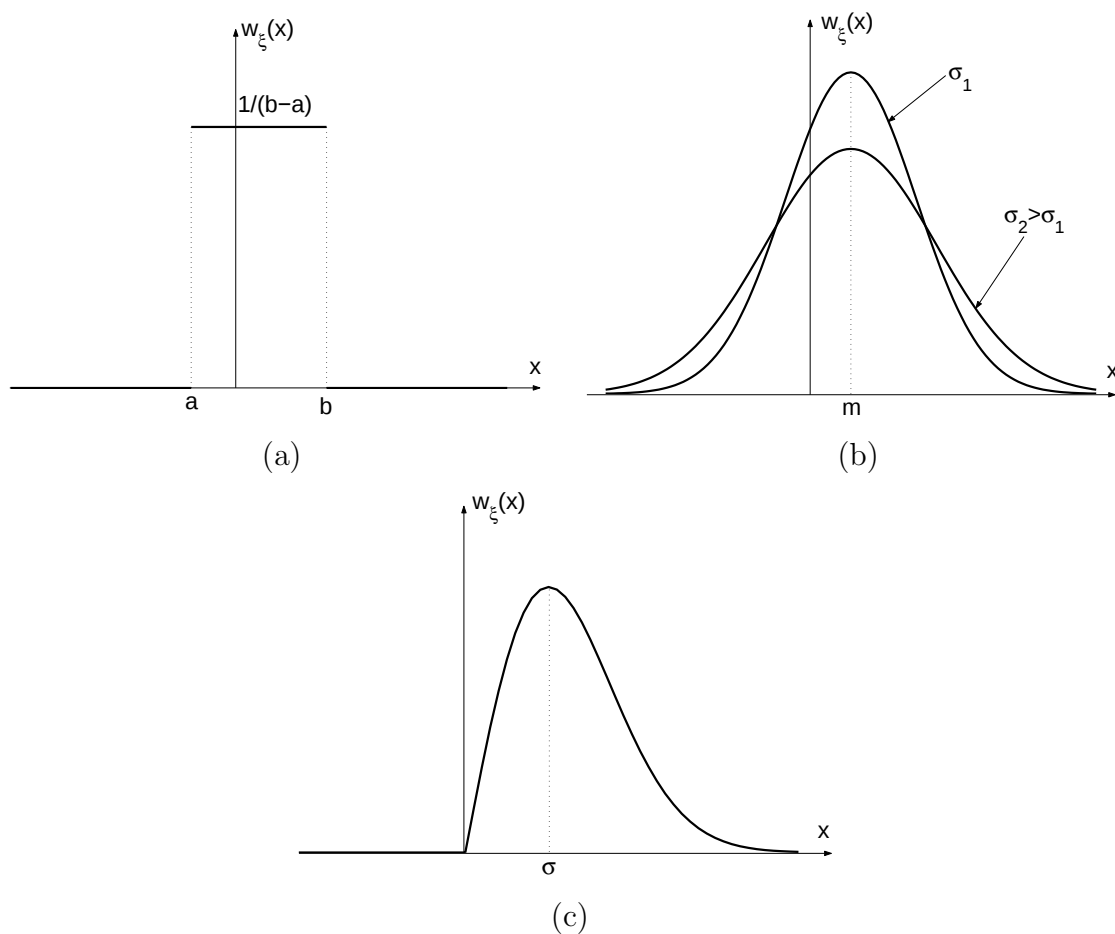


Figura 3.5: Tipuri de distribuții ale variabilelor aleatoare: (a) uniformă, (b) normală, (c) Rayleigh.

Graficul densității de probabilitate uniforme este prezentat în figura 3.5.(a). Forma distribuției indică faptul că o variabilă aleatoare uniformă poate lua cu aceeași probabilitate orice valoare în intervalul $[a, b]$, dar nu poate lua nici o valoare în exteriorul acestuia.

Prin calcul direct, se arată că media și dispersia distribuției uniforme sunt: $\bar{\xi} = \frac{a+b}{2}$ și $\sigma_{\xi}^2 = \frac{(b-a)^2}{12}$.

Distribuția uniformă modelează destule fenomene reale de interes (cum ar fi, spre exemplu, eroarea obținută la cuantizarea uniformă cu un număr suficient de mare de nivele).

3.7.2 Distribuția gaussiană (normală)

Se zice că o variabilă aleatoare este distribuită după o lege normală (sau gaussiană) de parametri $m \in \mathbb{R}$ și $\sigma > 0$ (și se notează $\xi: \mathcal{N}(m, \sigma)$) dacă densitatea ei de probabilitate este dată de:

$$w_\xi(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right). \quad (3.23)$$

Graficul distribuției gaussiene (cunoscut și sub numele de “clopotul lui Gauss”) este prezentat în figura 3.5.(b) pentru două valori diferite ale parametrului σ . Se observă că pe măsură de σ crește, clopotul se aplatizează și, în același timp, se lățește.

Prin calcul, se determină că media și dispersia distribuției normale sunt $\bar{\xi} = m$, respectiv $\sigma_\xi = \sigma$. Așadar, semnificația celor doi parametri de care depinde gaussiană este de medie și dispersie a variabilei aleatoare distribuite după o astfel de lege⁶.

Distribuția normală este cu siguranță cea mai importantă în studiul prelucrării statistice a semnalelor. Ipoteza de “normalitate” a distribuției multor fenomene reale este justificată de o teoremă (numită teorema limită centrală) despre care vom discuta mai pe larg în capitolul 4.

3.7.3 Distribuția Rayleigh

Distribuția Rayleigh este dată de relația:

$$w_\xi(x) = \begin{cases} \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right) & \text{dacă } x \geq 0 \\ 0 & \text{în rest.} \end{cases} \quad (3.24)$$

și este ilustrată în figura 3.5.(c).

Media și dispersia variabilei distribuite după o lege Rayleigh sunt $\bar{\xi} = \sigma\sqrt{\frac{\pi}{2}}$, respectiv $\sigma_\xi = \sigma\sqrt{\frac{4-\pi}{2}}$.

3.8 Funcții de o variabilă aleatoare

Fie ξ o variabilă aleatoare cu distribuție w_ξ cunoscută, și fie $g: \mathbb{R} \rightarrow \mathbb{R}$ o funcție de asemenea cunoscută. Aplicând funcția g variabilei ξ , obținem o altă variabilă aleatoare, pe care-o notăm $\eta = g(\xi)$. Problema pe care ne propunem să o rezolvăm în acest paragraf este calculul densității de probabilitate a variabilei aleatoare transformate w_η pe baza cunoștințelor pe care le avem despre distribuția variabilei inițiale w_ξ și despre funcția de transformare g .

⁶Se poate ilustra aici afirmația enunțată anterior despre semnificația dispersiei: cu cât σ crește, cu atât clopotul este mai “plat”, deci crește probabilitatea ca variabila să devieze mult în stânga și dreapta mediei. Invers, pe măsură ce σ scade, clopotul se îngustează, deci probabilitatea ca ξ să ia valori puternic deviate față de medie scade. S-a calculat că pentru distribuția gaussiană, probabilitatea ca variabila aleatoare să devieze cu mai mult de 3σ față de valoarea medie este de 0,3%: $P(m - 3\sigma \leq \xi \leq m + 3\sigma) = 99,7\%$ indiferent de σ .

Să ilustrăm la început problema pe un exemplu oarecare. Fie funcția $g(x)$ având graficul prezentat în figura 3.6. Variabila aleatoare ξ ia valori pe axa Ox , în timp ce $\eta = g(\xi)$ pe axa Oy .

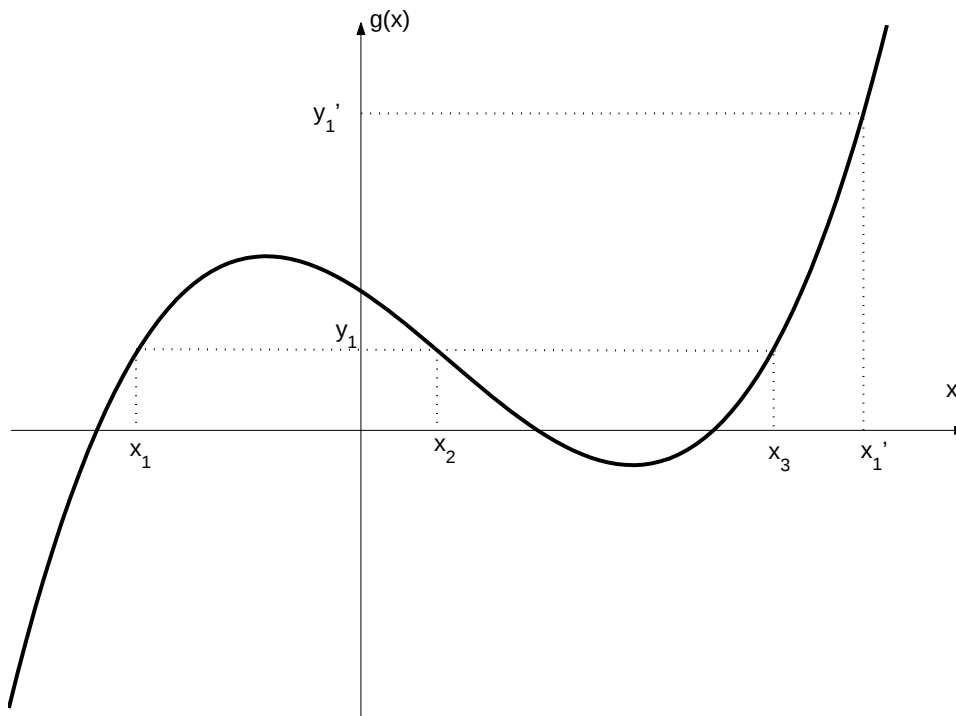


Figura 3.6: Exemplificarea problemei funcțiilor de o variabilă aleatoare.

Se observă, pentru valorile din figură că, spre exemplu:

$$\begin{aligned} F_{\eta}(y'_1) &= P(\eta \leq y'_1) = P(\xi \leq x'_1) = F_{\xi}(x'_1) \Big|_{g(x'_1)=y'_1} \\ F_{\eta}(y_1) &= P(\eta \leq y_1) = P((\xi \leq x_1) \cup (x_2 \leq \xi \leq x_3)) \\ &= F_{\xi}(x_3) - F_{\xi}(x_2) + F_{\xi}(x_1) \Big|_{g(x_1)=g(x_2)=g(x_3)=y_1} \end{aligned}$$

Astfel, se vede că pentru $\forall y \in \mathbb{R}$, valoarea funcției de repartiție a lui η în punctul respectiv $F_{\eta}(y)$ poate fi scrisă în funcție de valorile funcției de repartiție a variabilei inițiale F_{ξ} în punctele care sunt soluții ale ecuației $g(x) = y$.

În continuare vom demonstra următoarea teoremă:

Teoremă. Fie ξ o variabilă aleatoare cu distribuție w_{ξ} cunoscută, și fie $\eta = g(\xi)$, cu $g(x)$ o funcție cunoscută. Atunci, pentru $\forall y \in \mathbb{R}$ cu proprietatea că ecuația $g(x) = y$ are un număr finit sau cel mult numărabil de soluții, pe care le notăm cu $\{x_1, x_2, \dots\}$, are loc relația:

$$w_{\eta}(y) = \sum_k \frac{w_{\xi}(x_k)}{|g'(x_k)|} \quad (3.25)$$

Demonstrație. Să revenim la exemplul considerat anterior. Se observă din figura 3.7 că ecuația $g(x) = y_1$ admite trei soluții: $\{x_1, x_2, x_3\}$. Să considerăm în continuare valoarea $y_1 + dy_1$. Ecuația $g(x) = y_1 + dy_1$ admite și ea trei soluții, pe care le notăm $\{x_1 + dx_1, x_2 + dx_2, x_3 + dx_3\}$ (vezi figura 3.7).

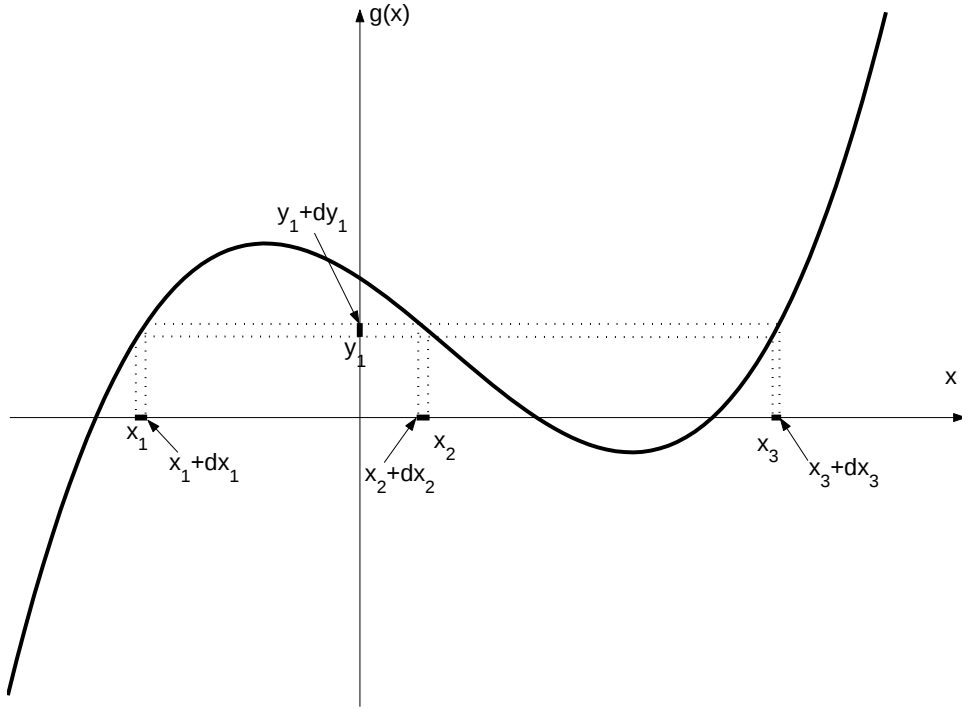


Figura 3.7: Ecuația $g(x) = y_1$ are soluțiile x_1, x_2 și x_3 .

În continuare, putem scrie:

$$P(y_1 < \eta \leq y_1 + dy_1) = P((x_1 < \xi \leq x_1 + dx_1) \cup (x_2 + dx_2 \leq \xi < x_2) \cup \dots \cup (x_3 < \xi \leq x_3 + dx_3)), \quad (3.26)$$

relație care este evidentă prin observarea figurii, unde intervalele care intervin sunt scoase în evidență. Se observă, de asemenea, că $dx_2 < 0$, întrucât x_2 se află pe o pantă negativă a funcției (adică $g'(x_2) < 0$).

Valoarea lui dy_1 poate fi făcută suficient de mică astfel încât intervalele $(x_i, x_i + dx_i]$ să nu se suprapună, oricât de apropiate ar fi x_i de x_j ⁷. În consecință, evenimentele $\{x_i < \xi \leq x_i + dx_i\}$ sunt incompatibile, și, aplicând axioma a treia a probabilităților,

⁷În această presupunere intervine necesitatea ca numărul soluțiilor ecuației $g(x) = y$ să fie finit, sau cel mult numărabil. În caz contrar, de exemplu, dacă soluția ecuației $g(x) = y$ ar fi un întreg interval $[a, b]$, atunci nu s-ar mai putea face ipoteza nesuprapunerii intervalelor $(x_i, x_i + dx_i]$ oricât de mic s-ar face dx_i .

avem:

$$P(y_1 < \eta \leq y_1 + dy_1) = P(x_1 < \xi \leq x_1 + dx_1) + P(x_2 + dx_2 \leq \xi < x_2) + \dots \\ \dots + P(x_3 < \xi \leq x_3 + dx_3). \quad (3.27)$$

Când $dy_1 \searrow$, putem aplica formula (3.6), și, în consecință, relația (3.27) devine:

$$w_\eta(y_1)dy_1 = w_\xi(x_1)dx_1 + w_\xi(x_2)|dx_2| + w_\xi(x_3)dx_3. \quad (3.28)$$

Este evidentă necesitatea de a-l lua pe dx_2 în modul în relația de mai sus, întrucât termenul respectiv trebuie să fie pozitiv (reprezintă o probabilitate).

Pe de altă parte, cum x_1, x_2 și x_3 sunt soluțiile ecuației $g(x) = y_1$, putem scrie:

$$y_1 = g(x_1) = g(x_2) = g(x_3), \quad (3.29)$$

de unde, prin diferențiere, obținem:

$$dy_1 = g'(x_1)dx_1 = g'(x_2)dx_2 = g'(x_3)dx_3. \quad (3.30)$$

Înlocuind dx_i cu $\frac{dy_1}{g'(x_i)}$ în ecuația (3.28) și simplificând pe dy_1 , obținem:

$$w_\eta(y_1) = \frac{w_\xi(x_1)}{g'(x_1)} + \frac{w_\xi(x_2)}{|g'(x_2)|} + \frac{w_\xi(x_3)}{g'(x_3)}. \quad (3.31)$$

care, ținând cont că atât $g'(x_1)$ cât și $g'(x_3)$ sunt pozitive, deci pot fi înlocuite cu modulul lor, reprezintă exact relația ce trebuia demonstrată!

Trebuie să observăm că raționamentul poate fi reluat în mod identic pentru orice formă a graficului funcției $g(x)$, ceea ce face ca demonstrația de mai sus să fie valabilă în cazul general.

Exemplu. Să se calculeze distribuția variabilei aleatoare $\eta = a\xi + b$, cu $a, b \in \mathbb{R}$, $a \neq 0$, în funcție de distribuția variabilei aleatoare ξ .

Rezolvare. Funcția de trecere între cele două variabile este funcția $g(x) = ax + b$. Pentru $\forall y \in \mathbb{R}$, ecuația $g(x) = y$ admite o singură soluție, și anume $x_1 = \frac{y-b}{a}$. Cum $g'(x) = a$, relația (3.25) se scrie:

$$w_\eta(y) = \frac{w_\xi(x_1)}{|g'(x_1)|} = \frac{1}{|a|} w_\xi\left(\frac{y-b}{a}\right). \quad (3.32)$$

În particular, dacă ξ are o distribuție uniformă sau gaussiană, rezultă din relația de mai sus că și η are același tip de distribuție, dar cu parametri modificați.

Exemplu. Să se calculeze distribuția variabilei aleatoare $\eta = \cos(\xi)$, unde ξ este o variabilă aleatoare distribuită uniform în intervalul $[0, 2\pi]$.

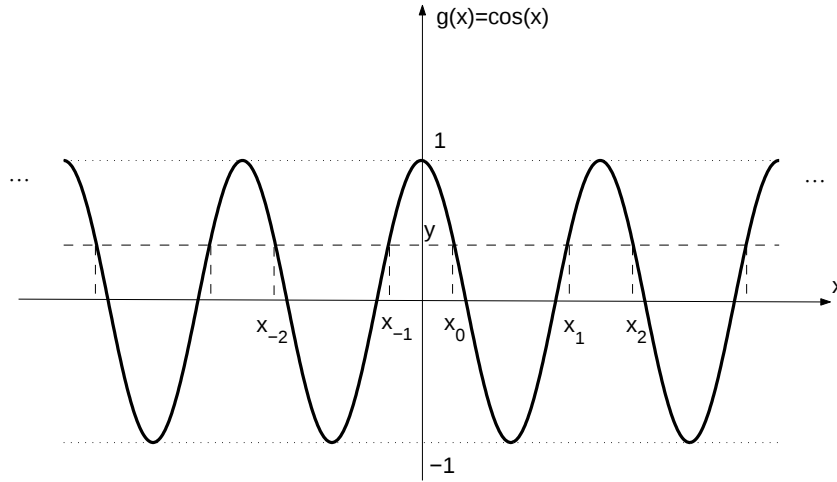


Figura 3.8: Funcția $g: \mathbb{R} \rightarrow [-1, 1]$ $g(x) = \cos(x)$.

Rezolvare. Graficul funcției g este prezentat în figura 3.8. Se observă că pentru $y \notin [-1, 1]$, ecuația $g(x) = y$ nu admite nici o soluție, și, deci, $w_\eta(y) = 0$.

Pentru $y \in [-1, 1]$, însă, ecuația $g(x) = y$ admite o infinitate numărabilă de soluții (v. figura 3.8), date de⁸:

$$x_k = \begin{cases} \arccos(y) + k\pi & \text{pentru } k = 2l \\ -\arccos(y) + (k+1)\pi & \text{pentru } k = 2l+1 \end{cases} \quad (3.33)$$

Observând, în plus, că $g'(x) = -\sin(x)$, relația (3.25) poate fi scrisă:

$$w_\eta(y) = \sum_{k \in \mathbb{Z}} \frac{w_\xi(x_k)}{|-\sin(x_k)|}. \quad (3.34)$$

Mai rămâne de calculat valoarea densității de probabilitate w_ξ în toate punctele x_k date de (3.33). Conform (3.22), distribuția lui ξ este dată de:

$$w_\xi(x) = \begin{cases} \frac{1}{2\pi} & \text{dacă } x \in [0, 2\pi] \\ 0 & \text{în rest} \end{cases} \quad (3.35)$$

Deci, trebuie să identificăm în continuare acele valori x_k din intervalul $[0, 2\pi]$, interval pe care valoarea funcției w_ξ este nenulă. Reamintindu-ne că funcția $\arccos: [-1, 1] \rightarrow [0, \pi]$, rezultă că nu există decât două valori x_k în intervalul căutat, respectiv $x_0 = \arccos(y)$ și $x_1 = 2\pi - \arccos(y)$. Având în vedere că $w_\xi(x_k) = 0$ pentru $k \notin \{0, 1\}$, relația (3.34) devine:

$$w_\eta(y) = \frac{w_\xi(x_0)}{|g'(x_0)|} + \frac{w_\xi(x_1)}{|g'(x_1)|} = \frac{w_\xi(\arccos(y))}{|-\sin(\arccos(y))|} + \frac{w_\xi(2\pi - \arccos(y))}{|-\sin(2\pi - \arccos(y))|}. \quad (3.36)$$

⁸Relația de mai jos poate fi scrisă mai compact, dar mai puțin riguros matematic, sub forma $x_k = \pm \arccos(y) + 2k\pi$, cu $k \in \mathbb{Z}$

Ținând, în plus, cont că $\sin(\arccos y) = \sqrt{1 - y^2}$, avem:

$$w_\eta(y) = \frac{\frac{1}{2\pi}}{\sqrt{1-y^2}} + \frac{\frac{1}{2\pi}}{\sqrt{1-y^2}} = \frac{1}{\pi\sqrt{1-y^2}}. \quad (3.37)$$

În concluzie, am arătat că:

$$w_\eta(y) = \begin{cases} \frac{1}{\pi\sqrt{1-y^2}} & \text{pentru } y \in [-1, 1] \\ 0 & \text{în rest} \end{cases}. \quad (3.38)$$

Graficul funcției (3.38) este dat în figura 3.9.

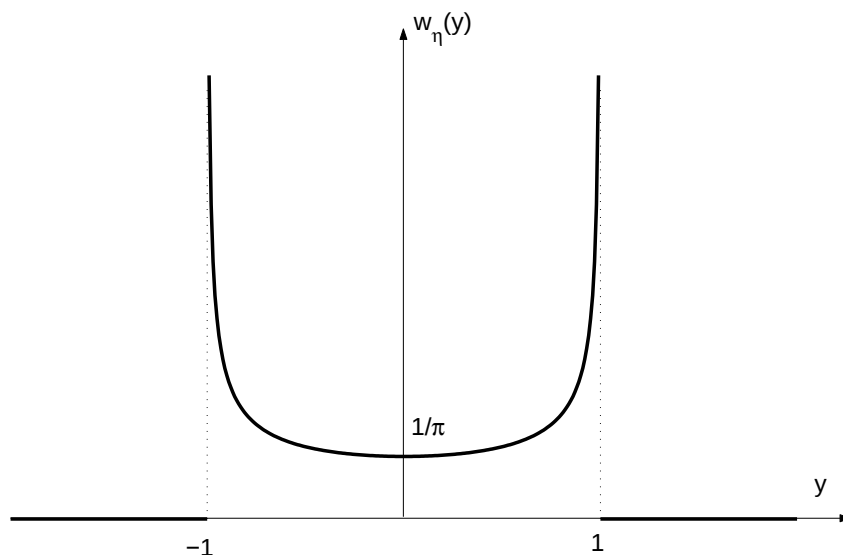


Figura 3.9: Graficul funcției $w_\eta(y)$ dată de relația 3.38.

Cazul soluțiilor nenumărabile.

În continuare, ne vom ocupa cu discuția cazului în care teorema (3.25) nu poate fi aplicată, respectiv cazul acelor $y_i \in \mathbb{R}$ pentru care ecuația $g(x) = y_i$ are ca soluție un întreg interval din $\mathbb{R}$: $\forall x \in [a, b], g(x) = y_i$.

Situația va fi discutată pe un exemplu. Fie $\eta = g(\xi)$, cu $g(x)$ dată de

$$g(x) = \begin{cases} -a & \text{dacă } x < -a \\ x & \text{dacă } -a \leq x \leq a \\ a & \text{dacă } x > a \end{cases}. \quad (3.39)$$

și ilustrată în figura 3.10. Presupunem că ξ are o distribuție $w_\xi(x)$ cunoscută, dar nespecificată.

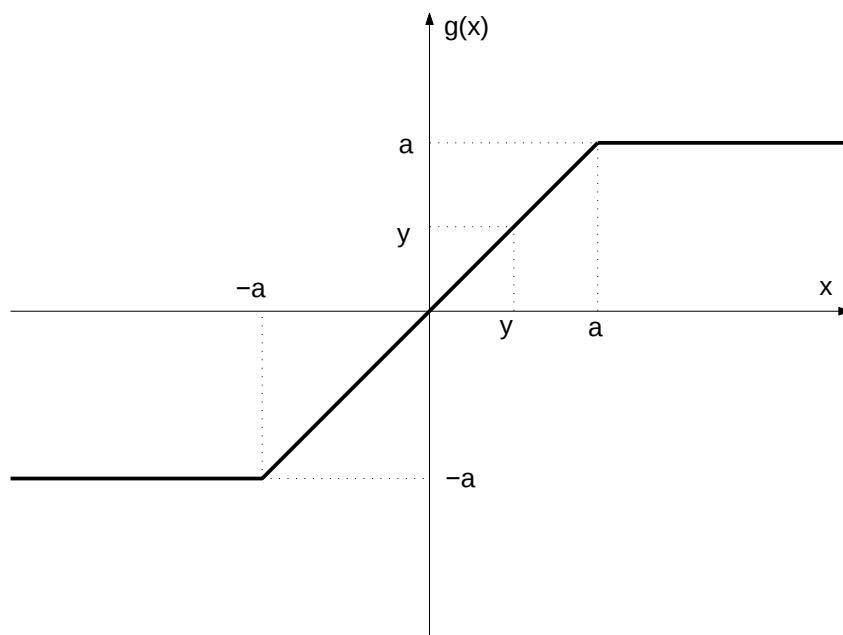


Figura 3.10: Funcția $g(x)$ dată de (3.39) .

Se observă că funcția g are două paliere, adică atât pentru $y_1 = -a$ cât și pentru $y_2 = a$, soluția ecuației $g(x) = y_1$ (respectiv $g(x) = y_2$) este intervalul $(-\infty, -a]$ (respectiv $[a, \infty)$) care sunt mulțimi nenumărabile.

Vom proceda, la fel ca și în exemplul dat la începutul acestui paragraf, prin a calcula valorile funcției de repartiție a variabilei transformate F_η .

Se poate constata cu ușurință urmărind graficul funcției g din figura 3.10 că:

$$\begin{aligned}
 y < -a, \quad F_\eta(y) &= P(\eta \leq y) = 0 \\
 y = -a, \quad F_\eta(-a) &= P(\eta \leq -a) = P(\xi \leq -a) = F_\xi(-a) \\
 y \in (-a, a), \quad F_\eta(y) &= P(\eta \leq y) = P(\xi \leq y) = F_\xi(y) \\
 y = a, \quad F_\eta(a) &= P(\eta \leq a) = P(\xi \leq \infty) = 1 \\
 y > a, \quad F_\eta(y) &= P(\eta \leq y) = P(\xi \leq \infty) = 1
 \end{aligned}$$

Sintetizând, avem:

$$F_\eta(y) = \begin{cases} 0 & \text{dacă } y < -a \\ F_\xi(y) & \text{dacă } -a \leq y < a \\ 1 & \text{dacă } y \geq a \end{cases} . \quad (3.40)$$

În coloana din stânga a figurii 3.11 sunt prezentate grafic funcția de repartiție a lui ξ (de o formă arbitrară) și cea a lui η . Scopul acestei reprezentări grafice este ilustrarea apariției unor discontinuități în funcția de repartiție a variabilei transformate F_η exact în punctele “speciale” $y_1 = -a$ și $y_2 = a$.

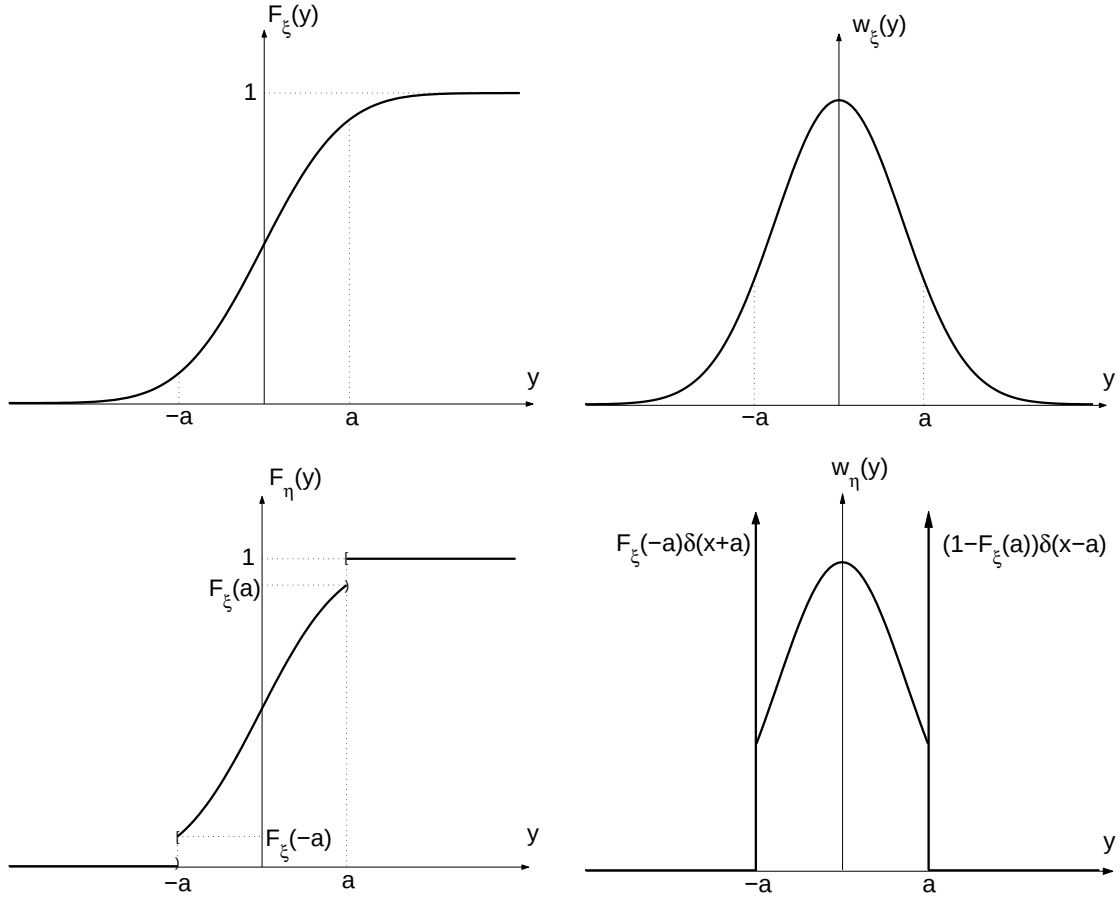


Figura 3.11: Funcțiile de repartiție și densitățile de probabilitate ale lui ξ și $\eta = g(x)$.

Aceste discontinuități din funcția de repartiție vor conduce, după cum am discutat în detaliu în paragraful 3.4, la apariția în densitatea de probabilitate w_η a unor impulsuri Dirac în punctele respective, de arie egală cu amplitudinea saltului făcut de F_η . Densitățile de probabilitate ale lui ξ și η sunt prezentate în coloana din dreapta din figura 3.11. Se observă că ariile celor două impulsuri Dirac sunt $F_\xi(-a) = P(\eta \leq -a) = P(\eta \in (-\infty, -a])$ pentru cel localizat în punctul $y_1 = -a$, respectiv $1 - F_\xi(a) = P(\eta \geq a) = P(\eta \in [a, \infty))$ pentru impulsul din $y_2 = a$. Din toate acestea, putem trage următoarea concluzie:

Pentru acei $y_i \in \mathbb{R}$ pentru care soluția ecuației $g(x) = y_i$ este un interval $[a, b]$, în densitatea de probabilitate a variabilei transformate w_η va apărea un impuls Dirac în punctul $y = y_i$ de arie egală cu probabilitatea $P(\xi \in [a, b]) = F_\xi(b) - F_\xi(a)$.

Cu alte cuvinte, devine nenulă $P(\eta = y_i)$, ceea ce e normal: întrucât $g(x) = y_i \forall x \in [a, b]$, atunci $\eta = y_i$ cu aceeași probabilitate cu care $\xi \in [a, b]$!

Exemplu. Să se arate că aplicând unei variabile aleatoare continue ξ oarecare însăși funcția ei de repartiție F_ξ , se obține o variabilă aleatoare uniformă în $[0, 1]$.

Rezolvare. Trebuie, deci, demonstrat că variabila aleatoare $\eta = F_\xi(\xi)$ are distribuție uniformă în $[0, 1]$ pentru *orice* formă a funcției de repartiție F_ξ (cu singura constrângere ca aceasta să fie continuă pe $\mathbb{R}$).

Deși, la o primă vedere, problema nu pare a fi rezolvabilă din lipsă de date suficiente (nu se cunoaște funcția de transformare g !!), totuși, ea admite soluție, iar demonstrația este extrem de utilă pentru înțelegerea mai profundă a problemei funcțiilor de variabile aleatoare.

Să începem prin a observa că, datorită faptului că $F_\xi: \mathbb{R} \rightarrow [0, 1]$, pentru $y \notin [0, 1]$, ecuația $F_\xi(x) = y$ nu are soluții reale, și, deci, $w_\eta(y) = 0$.

Pentru $y \in [0, 1]$, trebuie determinat numărul de soluții ale ecuației $F_\xi(x) = y$. Ne vom reaminti că funcția de repartiție a oricărei variabile aleatoare este crescătoare pe $\mathbb{R}$ (proprietatea 3). Dacă F_ξ ar fi strict crescătoare, problema ar fi rezolvată, pentru că funcțiile strict monotone sunt bijective, și, deci, ecuația $F_\xi(x) = y$ ar avea o singură soluție $\forall y \in [0, 1]$ și $\forall F_\xi$. Totuși, funcția este numai crescătoare, ceea ce înseamnă că poate admite porțiuni pe $\mathbb{R}$ în care “stagnează”, mai precis, $F_\xi(y)$ poate avea paliere, oricât de multe, și oricât de întinse! Rezolvarea s-ar opri aici, din lipsă de suficiente date, dacă nu am observa că un interval din $\mathbb{R}$ pe care F_ξ are palier este un interval *în care variabila ξ nu poate lua valori!* Într-adevăr:

$$F_\xi(x) = y_1 \quad \forall x \in [a, b] \Rightarrow P(\xi \in [a, b]) = F_\xi(b) - F_\xi(a) = y_1 - y_1 = 0.$$

Or, să nu uităm că noi aplicăm funcția F_ξ însăși variabilei aleatoare ξ . Rezultă, deci, că ne interesează numai restricția funcției F_ξ la intervalele în care ξ ia valori, adică intervalele pe care $w_\xi(x) > 0$, adică intervalele pe care funcția F_ξ este *strict crescătoare*. Figura 3.12 prezintă un exemplu.

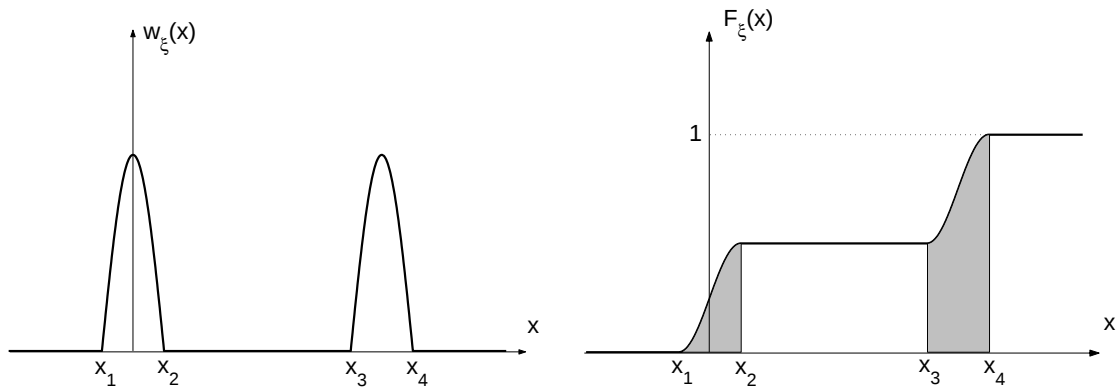


Figura 3.12: O funcție de densitate de probabilitate w_ξ arbitrară, și funcția de repartiție F_ξ corespunzătoare. Se observă că variabila aleatoare ia valori numai pe reuniunea de intervale $[x_1, x_2] \cup [x_3, x_4]$. Întrucât $\eta = F_\xi(\xi)$, de interes e numai restricția funcției de repartiție la intervalele respective (hașurate în figură): $F_\xi : ([x_1, x_2] \cup [x_3, x_4]) \rightarrow [0, 1]$. Astfel definită, funcția de repartiție e strict crescătoare.

Întrucât am decis asupra faptului că funcția de trecere este strict crescătoare, ea este bijectivă, deci $\forall y_1 \in [0, 1], \exists! x_1$ astfel încât $y_1 = F_\xi(x_1)$. Aplicând relația (3.25), avem:

$$w_\xi(y_1) = \frac{w_\xi(x_1)}{|F'_\xi(x_1)|} = \frac{w_\xi(x_1)}{|w_\xi(x_1)|} = \frac{w_\xi(x_1)}{w_\xi(x_1)} = 1.$$

În dezvoltarea de mai sus, am folosit faptul că derivata funcției de repartiție este însăși densitatea de probabilitate, și că aceasta este pozitivă, deci este egală cu modulul ei. Sintetizând, am arătat că:

$$w_\eta(y) = \begin{cases} 1 & \text{dacă } y \in [0, 1] \\ 0 & \text{în rest} \end{cases},$$

ceea ce termină demonstrația noastră.

3.9 Teorema de medie

Teorema de medie pe care o vom enunța și demonstra în acest paragraf permite calculul mediei variabilei aleatoare transformate fără a-i mai calcula acesteia distribuția, ci direct din distribuția variabilei inițiale și din funcția de trecere dintre cele două variabile aleatoare.

Teoremă. Fie $\eta = g(\xi)$. Atunci, media lui η poate fi calculată ca:

$$\bar{\eta} = \int_{-\infty}^{\infty} g(x)w_\xi(x)dx. \quad (3.41)$$

Demonstrație. Să revenim la situația ilustrată în figura 3.7. Înmulțind relația (3.28) cu y_1 , și ținând cont de (3.29), avem:

$$y_1 w_\eta(y_1) dy_1 = g(x_1)w_\xi(x_1)dx_1 + g(x_2)w_\xi(x_2)|dx_2| + g(x_3)w_\xi(x_3)dx_3. \quad (3.42)$$

Valoarea medie a lui η poate fi scrisă:

$$\bar{\eta} \approx \sum_{\substack{j \\ \cup dy_j = \mathbb{R}}} y_j w_\eta(y_j) dy_j \xrightarrow{dy_j \rightarrow 0} \int_{-\infty}^{\infty} y w_\eta(y) dy, \quad (3.43)$$

cu alte cuvinte, valoarea $\bar{\eta}$ se obține sumând toate cantitățile de tip $y_j w_\eta(y_j) dy_j$ pentru dy_j baleind întreaga axă reală Oy . Dar suma tuturor cantităților $y_j w_\eta(y_j) dy_j$ poate fi scrisă, conform relației (3.42) cu o sumă de cantități de tip $g(x_i)w_\xi(x_i)dx_i$. Întrucât g este o funcție, care, prin definiție, alocă *fiecărui x un y și numai unul*, putem afirma că în suma respectivă de cantități de tip $g(x_i)w_\xi(x_i)dx_i$, fiecare va apărea o dată și numai o dată. Cu alte cuvinte, când dy_j va baleia axa Oy , dx_i va baleia axa Ox ! Chiar dacă

unei cantități $y_j w_\eta(y_j) dy_j$ îi pot corepunde nici una, una sau mai multe cantități de tipul $g(x_i) w_\xi(x_i) dx_i$, invers, putem spune că unei cantități $g(x_i) w_\xi(x_i) dx_i$ îi va corespunde o singură cantitate $y_i w_\eta(y_i) dy_i$ și numai una! Deci, putem scrie:

$$\sum_{\substack{j \\ \cup_j dy_j = \mathbb{R}}} y_j w_\eta(y_j) dy_j \approx \sum_{\substack{i \\ \cup_i dx_i = \mathbb{R}}} g(x_i) w_\xi(x_i) dx_i, \quad (3.44)$$

de unde, prin trecere la limită:

$$\bar{\eta} = \int_{-\infty}^{\infty} y w_\eta(y) dy = \int_{-\infty}^{\infty} g(x) w_\xi(x) dx. \quad (3.45)$$

Observație. Deși demonstrația teoremei s-a făcut în ipoteza particulară a unui număr cel mult numărabil de soluții ale ecuației $g(x) = y$, relația (3.41) este valabilă pentru orice funcție $g(x)$.

Exemplu. Să se calculeze media lui $\eta = a\xi + b$ cu $a, b \in \mathbb{R}$. Conform teoremei de medie, putem scrie:

$$\bar{\eta} = \int_{-\infty}^{\infty} (ax + b) w_\xi(x) dx = a \underbrace{\int_{-\infty}^{\infty} x w_\xi(x) dx}_{\bar{\xi}} + b \underbrace{\int_{-\infty}^{\infty} w_\xi(x) dx}_1 = a\bar{\xi} + b. \quad (3.46)$$

Spre exemplu, pentru $b = 0$, se poate scrie:

$$\overline{a\xi} = a\bar{\xi}, \quad (3.47)$$

ceea ce se interpretează în sensul că înmulțirea cu o constantă comută cu medierea statistică (altfel spus, constantele “ies în afara” operației de mediere). De-a lungul acestei lucrări, vom face de mai multe ori apel la acest rezultat.

Capitolul 4

Caracterizarea unei perechi de variabile aleatoare

4.1 Introducere

În capitolul precedent, am discutat despre caracterizarea statistică a unei variabile aleatoare. Dacă avem două variabile aleatoare, caracterizarea fiecăreia dintre ele în parte cu mărimile mai sus menționate permite descrierea completă a comportamentului variabilei respective *independent de cealaltă variabilă*. Cu alte cuvinte, nu avem suficientă informație pentru a putea evalua eventuala interdependență statistică între cele două variabile considerate. Această interdependență poate fi studiată numai folosind mărimi statistice de ordinul doi, specifice *perechii* de variabile aleatoare. Vom arăta că această caracterizare de ordinul doi a unei perechi de variabile aleatoare este suficient de completă, ea înglobând, ca un caz particular, caracterizarea de ordinul unu a fiecăreia dintre cele două variabile aleatoare.

4.2 Funcția de repartiție de ordinul doi

Fie ξ și η două variabile aleatoare. Se definește funcția de repartiție de ordinul doi a perechii de variabile aleatoare ca fiind $F_{\xi\eta}: \mathbb{R}^2 \rightarrow [0, 1]$, cu expresia dată de:

$$F_{\xi\eta}(x, y) = P((\xi \leq x) \cap (\eta \leq y)). \quad (4.1)$$

Cu alte cuvinte, valoarea funcției de repartiție de ordinul doi într-un punct $(x_0, y_0) \in \mathbb{R}^2$ reprezintă probabilitatea ca, în același timp, ξ să fie mai mic decât x_0 și η mai mic decât y_0 ; altfel spus, probabilitatea ca punctul (ξ, η) să se afle în domeniul $D \in \mathbb{R}^2$ figurat în figura 4.1.

Proprietățile funcției de repartiție de ordinul doi

1. $F_{\xi\eta}(-\infty, y) = F_{\xi\eta}(x, -\infty) = 0$.

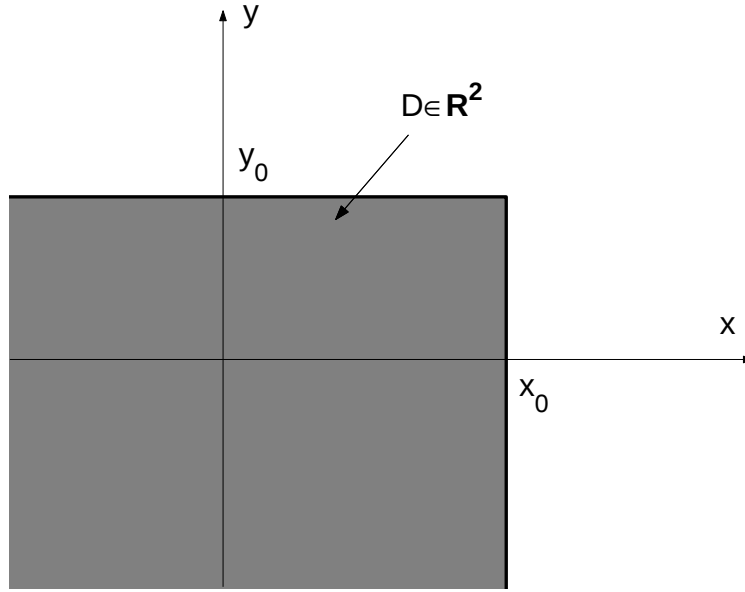


Figura 4.1: Domeniul de valori al punctului (ξ, η) .

Această proprietate este evidentă, având în vedere că evenimentele $\xi \leq -\infty$, respectiv $\eta \leq -\infty$ sunt de probabilitate zero.

2. $F_{\xi\eta}(\infty, \infty) = 1$.

Iarăși, o proprietate evidentă: conform definiției, $F_{\xi\eta}(\infty, \infty)$ reprezintă probabilitatea ca ξ și η să ia concomitent valori în $\mathbb{R}$, ceea ce reprezintă evenimentul sigur.

3. $F_{\xi\eta}(x, \infty) = F_{\xi}(x)$, respectiv $F_{\xi\eta}(\infty, y) = F_{\eta}(y)$.

Demonstrație. $F_{\xi\eta}(x, \infty) = P((\xi \leq x) \cap (\eta \leq \infty)) = P(\xi \leq x) = F_{\xi}(x)$.

Această proprietate este foarte importantă, întrucât demonstrează ceea ce am enunțat în introducere, și anume faptul că mărimile statistice de ordinul unu specifice fiecăreia dintre variabilele aleatoare pot fi obținute ca un caz particular al mărimilor de ordinul doi ale perechii de variabile aleatoare.

4. $P((x_1 < \xi \leq x_2) \cap (y_1 < \eta \leq y_2)) = F_{\xi\eta}(x_2, y_2) - F_{\xi\eta}(x_1, y_2) - F_{\xi\eta}(x_2, y_1) + F_{\xi\eta}(x_1, y_1)$.

Demonstrație. Avem

$$\begin{aligned} P((x_1 < \xi \leq x_2) \cap (y_1 < \eta \leq y_2)) &= \\ &= P((x_1 < \xi \leq x_2) \cap (\eta \leq y_2)) - P((x_1 < \xi \leq x_2) \cap (\eta \leq y_1)) \end{aligned}$$

Apoi, observăm că fiecare dintre cei doi termeni de mai sus poate fi la rândul său

descompus:

$$\begin{aligned} P((x_1 < \xi \leq x_2) \cap (\eta \leq y_2)) &= P((\xi \leq x_2) \cap (\eta \leq y_2)) - P((x_1 < \xi) \cap (\eta \leq y_2)) \\ &= F_{\xi\eta}(x_2, y_2) - F_{\xi\eta}(x_1, y_2), \end{aligned}$$

respectiv

$$\begin{aligned} P((x_1 < \xi \leq x_2) \cap (\eta \leq y_1)) &= P((\xi \leq x_2) \cap (\eta \leq y_1)) - P((x_1 < \xi) \cap (\eta \leq y_1)) \\ &= F_{\xi\eta}(x_2, y_1) - F_{\xi\eta}(x_1, y_1). \end{aligned}$$

Apoi, prin înlocuirea relațiilor de mai sus, obținem rezultatul ce trebuia demonstrat.

4.3 Densitatea de probabilitate de ordinul doi

Se definește densitatea de probabilitate de ordinul doi a perechii de variabile aleatoare (ξ, η) ca:

$$w_{\xi\eta}(x, y) = \frac{\partial^2 F_{\xi\eta}(x, y)}{\partial x \partial y}. \quad (4.2)$$

Explicitând formula derivatei, putem scrie:

$$\frac{\partial F_{\xi\eta}(x, y)}{\partial x} = \lim_{\Delta x \rightarrow 0} \frac{F_{\xi\eta}(x + \Delta x, y) - F_{\xi\eta}(x, y)}{\Delta x},$$

după care:

$$\begin{aligned} w_{\xi\eta}(x, y) &= \frac{\partial^2 F_{\xi\eta}(x, y)}{\partial x \partial y} = \frac{\partial \left(\frac{\partial F_{\xi\eta}(x, y)}{\partial x} \right)}{\partial y} = \frac{\partial \left(\lim_{\Delta x \rightarrow 0} \frac{F_{\xi\eta}(x + \Delta x, y) - F_{\xi\eta}(x, y)}{\Delta x} \right)}{\partial y} \\ &= \lim_{\Delta x \rightarrow 0} \frac{\frac{\partial F_{\xi\eta}(x + \Delta x, y)}{\partial y} - \frac{\partial F_{\xi\eta}(x, y)}{\partial y}}{\Delta x} \\ &= \lim_{\Delta x \rightarrow 0} \frac{\lim_{\Delta y \rightarrow 0} \frac{F_{\xi\eta}(x + \Delta x, y + \Delta y) - F_{\xi\eta}(x + \Delta x, y)}{\Delta y} - \lim_{\Delta y \rightarrow 0} \frac{F_{\xi\eta}(x, y + \Delta y) - F_{\xi\eta}(x, y)}{\Delta y}}{\Delta x} \\ &= \lim_{\Delta x \rightarrow 0} \lim_{\Delta y \rightarrow 0} \frac{F_{\xi\eta}(x + \Delta x, y + \Delta y) - F_{\xi\eta}(x + \Delta x, y) - F_{\xi\eta}(x, y + \Delta y) + F_{\xi\eta}(x, y)}{\Delta x \Delta y}, \end{aligned}$$

adică, ținând cont de proprietatea 4 a funcției de repartiție de ordinul doi, putem scrie:

$$w_{\xi\eta}(x, y) = \lim_{\Delta x \rightarrow 0} \lim_{\Delta y \rightarrow 0} \frac{P((\xi \in (x, x + \Delta x]) \cap (\eta \in (y, y + \Delta y])))}{\Delta x \Delta y}. \quad (4.3)$$

Relația (4.3) poate fi rescrisă ca:

$$w_{\xi\eta}(x, y) \Delta x \Delta y \underset{\Delta x, \Delta y \searrow 0}{\approx} P((\xi \in (x, x + \Delta x]) \cap (\eta \in (y, y + \Delta y])). \quad (4.4)$$

Relația (4.4) poate fi interpretată la fel ca și în cazul unidimensional: cu ajutorul densității de probabilitate de ordinul doi $w_{\xi\eta}$ se poate calcula probabilitatea ca, în același timp, ξ să se afle în intervalul infinitesimal $(x, x + \Delta x]$ iar η să se afle în intervalul $(y, y + \Delta y]$, iar această probabilitate este dată de *volumul* de sub graficul funcției densitate de probabilitate delimitat de domeniul $(x, x + \Delta x] \times (y, y + \Delta y]$.

Proprietățile densității de probabilitate de ordinul doi

1. $w_{\xi\eta}(x, y) \geq 0, \forall (x, y) \in \mathbb{R}^2$.
2. Pentru orice domeniu $D \in \mathbb{R}^2$ care reprezintă un eveniment, $P((\xi, \eta) \in D) = \iint_D w_{\xi\eta}(x, y) dx dy$.

Demonstrație. Orice domeniu plan D poate fi descompus ca o reuniune de subdomenii infinitezimale dreptunghiulare nesuprapuse $(x_i, x_i + \Delta x_i] \times (y_j, y_j + \Delta y_j]$, după cum este arătat în figura 4.2.

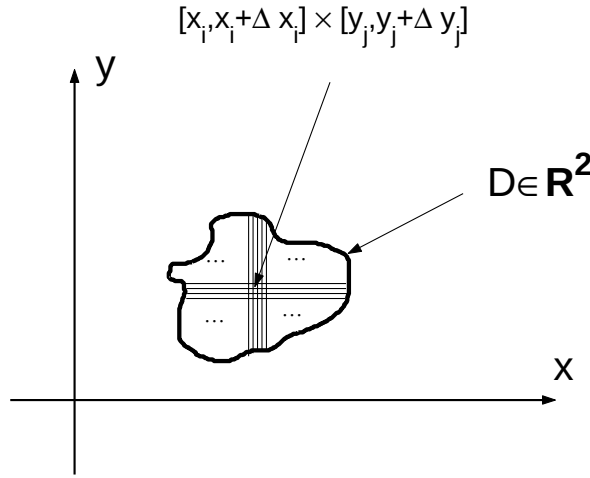


Figura 4.2: Descompunerea unui domeniu plan într-o reuniune de subdomenii infinitezimale nesuprapuse.

Atunci, se poate scrie:

$$\begin{aligned} P((\xi, \eta) \in D) &\approx P\left((\xi, \eta) \in \bigcup_i \bigcup_j ([x_i, x_i + \Delta x_i] \times [y_j, y_j + \Delta y_j])\right) \\ &= \sum_i \sum_j P((\xi \in [x_i, x_i + \Delta x_i]) \cap (\eta \in [y_j, y_j + \Delta y_j])), \end{aligned} \quad (4.5)$$

ceea ce, ținând cont de (4.4), devine:

$$P((\xi, \eta) \in D) \approx \sum_i \sum_j w_{\xi\eta}(x_i, y_j) \Delta x_i \Delta y_j.$$

Făcând $\Delta x_i \rightarrow 0$ și $\Delta y_j \rightarrow 0$ în relația de mai sus, se obține, evident, relația pe care trebuia să o demonstrăm.

3. Condiția de normare a densității de probabilitate de ordinul doi:

$$\iint_{\mathbb{R}^2} w_{\xi\eta}(x, y) dx dy = 1$$

Demonstrație. Se scrie proprietatea 2 pentru $D = \mathbb{R}^2$.

$$4. F_{\xi\eta}(x, y) = \int_{-\infty}^x \int_{-\infty}^y w_{\xi\eta}(u, v) du dv.$$

Demonstrație. Se scrie proprietatea 2 pentru $D = (-\infty, x] \times (-\infty, y]$ și se ține cont de (4.1).

5. Calculul densităților de probabilitate de ordinul unu (numite, în acest context, densități *marginale*), în funcție de densitatea de probabilitate de ordinul doi:

$$w_{\xi}(x) = \int_{-\infty}^{\infty} w_{\xi\eta}(x, y) dy \quad (4.6a)$$

$$w_{\eta}(y) = \int_{-\infty}^{\infty} w_{\xi\eta}(x, y) dx \quad (4.6b)$$

Demonstrație. Vom demonstra numai prima dintre relațiile de mai sus, cealaltă demonstrându-se analog. Se scrie proprietatea 4 a densității de probabilitate pentru $y = \infty$, și, ținând cont de proprietatea 3 a funcției de repartiție, rezultă:

$$F_{\xi}(x) = F_{\xi\eta}(x, \infty) = \int_{-\infty}^x \int_{-\infty}^{\infty} w_{\xi\eta}(u, v) du dv, \quad (4.7)$$

de unde:

$$w_{\xi}(x) = \frac{dF_{\xi}(x)}{dx} = \frac{d}{dx} \int_{-\infty}^x \int_{-\infty}^{\infty} w_{\xi\eta}(u, v) du dv = \int_{-\infty}^{\infty} w_{\xi\eta}(x, v) dv. \quad (4.8)$$

În deducerea relației de mai sus s-a ținut cont de faptul că, pentru o funcție oarecare $f(x)$, are loc relația

$$\frac{d}{dx} \int_a^x f(u) du = f(x). \quad (4.9)$$

Această proprietate ne arată clar cât de multă informație suplimentară ne aduce caracterizarea de ordinul doi a perechii de variabile aleatoare față de caracterizarea de ordinul unu a fiecăreia dintre variabile. Practic, interpretarea relațiilor (4.6a) și (4.6b) este aceea că densitățile de probabilitate marginale se obțin prin *proiecția* funcției densitate de probabilitate de ordinul doi pe fiecare dintre cele două axe. Este clar că două funcții unidimensionale obținute prin proiecție nu pot conține la fel de multă informație ca funcția bidimensională care a fost proiectată. Rezultă că, în general, *nu se poate determina densitatea de probabilitate de ordinul doi pornind de la cele două distribuții de ordinul unu!*

4.4 Distribuții condiționate

În foarte multe situații din practică, este de interes cunoașterea distribuției lui η atunci când ξ ia o anumită valoare x , respectiv $w_{\eta|\{\xi=x\}}(y)$. Aplicarea directă a definiției (3.8) nu este posibilă, întrucât în general, după cum am arătat în capitolul 3, $P(\xi = x) = 0$. Totuși, densitatea respectivă poate fi definită printr-o trecere la limită. Să considerăm mai întâi evenimentul $A = \{x_1 < \xi \leq x_2\}$. Conform (3.7), avem:

$$F_{\eta|\{x_1 < \xi \leq x_2\}}(y) = \frac{P((\eta \leq y) \cap (x_1 < \xi \leq x_2))}{P(x_1 < \xi \leq x_2)} = \frac{F_{\xi\eta}(x_2, y) - F_{\xi\eta}(x_1, y)}{F_{\xi}(x_2) - F_{\xi}(x_1)}. \quad (4.10)$$

În relația de mai sus, am aplicat proprietatea 4 a funcției de repartiție de ordinul doi. Prin derivare după y , obținem:

$$w_{\eta|\{x_1 < \xi \leq x_2\}}(y) = \frac{dF_{\eta|\{x_1 < \xi \leq x_2\}}(y)}{dy} = \frac{\frac{\partial F_{\xi\eta}(x_2, y)}{\partial y} - \frac{\partial F_{\xi\eta}(x_1, y)}{\partial y}}{F_{\xi}(x_2) - F_{\xi}(x_1)}. \quad (4.11)$$

Ținând cont de proprietatea 4 a densității de probabilitate de ordinul doi și de relația (4.9), putem scrie:

$$\frac{\partial F_{\xi\eta}(x, y)}{\partial y} = \frac{\partial \int_{-\infty}^x \int_{-\infty}^y w_{\xi\eta}(u, v) du dv}{\partial y} = \int_{-\infty}^x w_{\xi\eta}(u, y) du, \quad (4.12)$$

de unde (4.11) se scrie:

$$w_{\eta|\{x_1 < \xi \leq x_2\}}(y) = \frac{\int_{-\infty}^{x_2} w_{\xi\eta}(u, y) du - \int_{-\infty}^{x_1} w_{\xi\eta}(u, y) du}{F_{\xi}(x_2) - F_{\xi}(x_1)} = \frac{\int_{x_1}^{x_2} w_{\xi\eta}(u, y) du}{F_{\xi}(x_2) - F_{\xi}(x_1)}. \quad (4.13)$$

Alegând $x_1 = x$, respectiv $x_2 = x + \Delta x$, putem rescrie pe (4.13) sub forma:

$$w_{\eta|\{x < \xi \leq x + \Delta x\}}(y) = \frac{\int_x^{x+\Delta x} w_{\xi\eta}(u, y) du}{F_{\xi}(x + \Delta x) - F_{\xi}(x)} \underset{\Delta x \searrow 0}{\approx} \frac{w_{\xi\eta}(x, y) \Delta x}{w_{\xi}(x) \Delta x}. \quad (4.14)$$

Putem, deci, scrie că:

$$w_{\eta|\{\xi=x\}}(y) = \frac{w_{\xi\eta}(x, y)}{w_{\xi}(x)}. \quad (4.15)$$

Pentru un x_1 fixat, funcția $w_{\xi\eta}(x_1, y)$ este o curbă ce reprezintă intersecția între suprafața $w_{\xi\eta}(x, y)$ cu planul $x = x_1$ (după cum este ilustrat în figura 4.3). Interpretarea relației (4.15) este că densitatea de probabilitate $w_{\eta|\{\xi=x_1\}}(y)$ reprezintă ecuația curbei respective normalizate cu $w_{\xi}(x_1)$, astfel încât să devină de arie unitară.

Evident, relația (4.15) poate fi scrisă în mod similar și pentru cealaltă variabilă aleatoare. Putem, deci, scrie că:

$$w_{\xi\eta}(x, y) = w_{\xi|\{\eta=y\}}(x)w_{\eta}(y) = w_{\eta|\{\xi=x\}}(y)w_{\xi}(x). \quad (4.16)$$

interpretarea funcției $w_{\xi|\{\eta=y\}}$ fiind de intersecție a suprafeței $w_{\xi\eta}(x, y)$ cu planul $y = \text{constant}$.

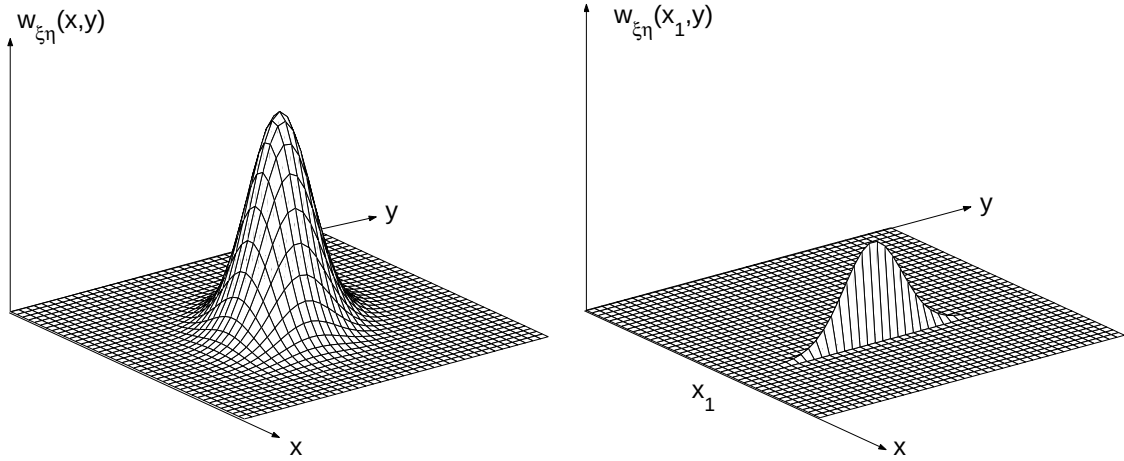


Figura 4.3: O densitate de probabilitate de ordinul doi și intersecția ei cu planul $x = x_1$.

4.5 Variabile aleatoare independente

Prin definiție, două variabile aleatoare ξ și η sunt independente dacă evenimentele $\{\xi \leq x\}$ și $\{\eta \leq y\}$ sunt independente pentru orice x și y în $\mathbb{R}$. Reamintindu-ne definiția a două evenimente independente (2.5), avem:

$$F_{\xi\eta}(x, y) = P((\xi \leq x) \cap (\eta \leq y)) = P(\xi \leq x)P(\eta \leq y) = F_{\xi}(x)F_{\eta}(y), \quad (4.17)$$

de unde, prin derivare după x și y , obținem:

$$w_{\xi\eta}(x, y) = w_{\xi}(x)w_{\eta}(y). \quad (4.18)$$

Evident, în ipoteza de independență, relațiile (4.16) și (4.15) devin:

$$w_{\xi|\{\eta=y\}}(x) = w_{\xi}(x), \quad (4.19)$$

$$w_{\eta|\{\xi=x\}}(y) = w_{\eta}(y). \quad (4.20)$$

Trebuie remarcat faptul că independența este singura ipoteză în care cunoașterea celor două distribuții de ordinul unu este suficientă pentru determinarea distribuției de ordinul doi a perechii de variabile aleatoare.

4.6 O funcție de două variabile aleatoare. Teorema limită centrală.

Presupunem că avem două variabile aleatoare ξ și η , având distribuția de ordinul doi $w_{\xi\eta}$ cunoscută. Obținem variabila aleatoare ζ aplicând funcția $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ (cunoscută) variabilelor aleatoare ξ și η :

$$\zeta = g(\xi, \eta). \quad (4.21)$$

La fel ca și în capitolul 3, ne punem problema calculului densității de probabilitate w_ζ a variabilei aleatoare transformate ζ în funcție de distribuția de ordinul doi a variabilelor aleatoare inițiale $w_{\xi\eta}$ și de funcția de trecere g . Avem următoarele relații:

$$F_\zeta(z) = P(\zeta \leq z) = P((\xi, \eta) \in D_z), \quad (4.22)$$

cu

$$D_z = \left\{ (x, y) \in \mathbb{R}^2 \mid g(x, y) \leq z \right\}. \quad (4.23)$$

Trebuie, deci, determinat domeniul D_z din (4.23), după care relația (4.22) poate fi scrisă (ținând cont de proprietatea 2 a densității de probabilitate de ordinul doi):

$$F_\zeta(z) = \iint_{D_z} w_{\xi\eta}(x, y) dx dy, \quad (4.24)$$

iar apoi

$$w_\zeta(z) = \frac{dF_\zeta(z)}{dz} = \frac{d}{dz} \iint_{D_z} w_{\xi\eta}(x, y) dx dy. \quad (4.25)$$

Problema care rămâne de rezolvat este determinarea domeniului D_z pentru *fiecare* $z \in \mathbb{R}$, operațiune care necesită o tratare separată pentru fiecare funcție g considerată. În cele ce urmează, vom da un exemplu de aplicare a celor descrise mai sus.

Exemplu. Distribuția sumei a două variabile aleatoare.

Considerăm funcția $g(x, y) = x + y$, ceea ce este echivalent cu a spune că:

$$\zeta = \xi + \eta. \quad (4.26)$$

Domeniul $D_z = \left\{ (x, y) \in \mathbb{R}^2 \mid x + y \leq z \right\}$ este cel prezentat în figura 4.4.

În aceste condiții, ecuația (4.24) devine:

$$F_\zeta(z) = \iint_{D_z} w_{\xi\eta}(x, y) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} w_{\xi\eta}(x, y) dx dy, \quad (4.27)$$

iar ecuația (4.25) devine, la rândul ei:

$$\begin{aligned} w_\zeta(z) &= \frac{d}{dz} \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} w_{\xi\eta}(x, y) dx dy = \int_{-\infty}^{\infty} \left(\frac{d}{dz} \int_{-\infty}^{z-y} w_{\xi\eta}(x, y) dx \right) dy \\ &= \int_{-\infty}^{\infty} w_{\xi\eta}(z - y, y) dy. \end{aligned} \quad (4.28)$$

În calculul relației (4.28) s-a ținut cont de (4.9).

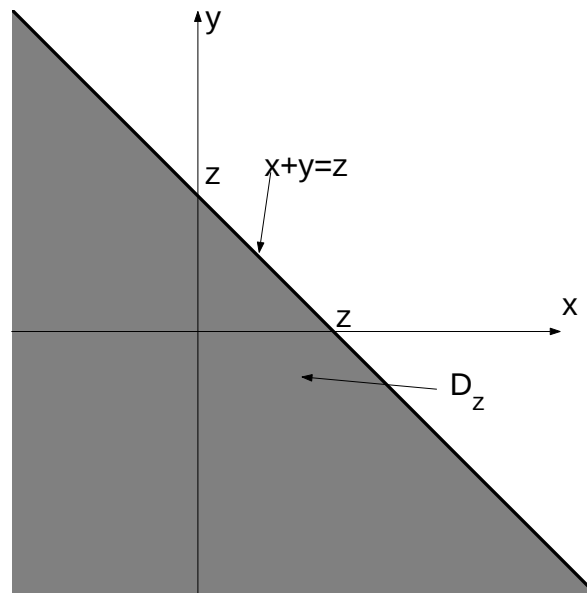


Figura 4.4: Domeniul D_z pentru funcția $g(x, y) = x + y$.

Dacă, în plus, ξ și η sunt independente, atunci relația (4.28) se scrie:

$$w_\zeta(z) = \int_{-\infty}^{\infty} w_\xi(z - y)w_\eta(y)dy = w_\xi(\zeta) \star w_\eta(\zeta). \quad (4.29)$$

Rezultatul de mai sus, care afirmă că densitatea de probabilitate a sumei a două variabile aleatoare independente se obține prin convoluția celor două distribuții, este foarte important, întrucât stă la baza demonstrației unei teoreme fundamentale în statistică, și anume *teorema limită centrală* care se enunță astfel:

Teoremă. Fie $\xi_1, \xi_2, \dots, \xi_N$ variabile aleatoare având distribuții *oarecare*, satisfăcând doar următoarele condiții:

1. oricare două dintre variabile, ξ_i și ξ_j , sunt independente,
2. nici una dintre variabile nu are o valoare preponderentă față de celelalte, ceea ce e echivalent cu a spune că varianța fiecăreia dintre ele este neglijabilă în raport cu suma varianțelor celorlalte:

$$\sigma_i^2 \ll \sum_{\substack{i=1 \\ i \neq j}}^N \sigma_j^2.$$

Atunci, variabila aleatoare $\eta = \xi_1 + \xi_2 + \dots + \xi_N$ are o distribuție gaussiană atunci când $N \rightarrow \infty$.

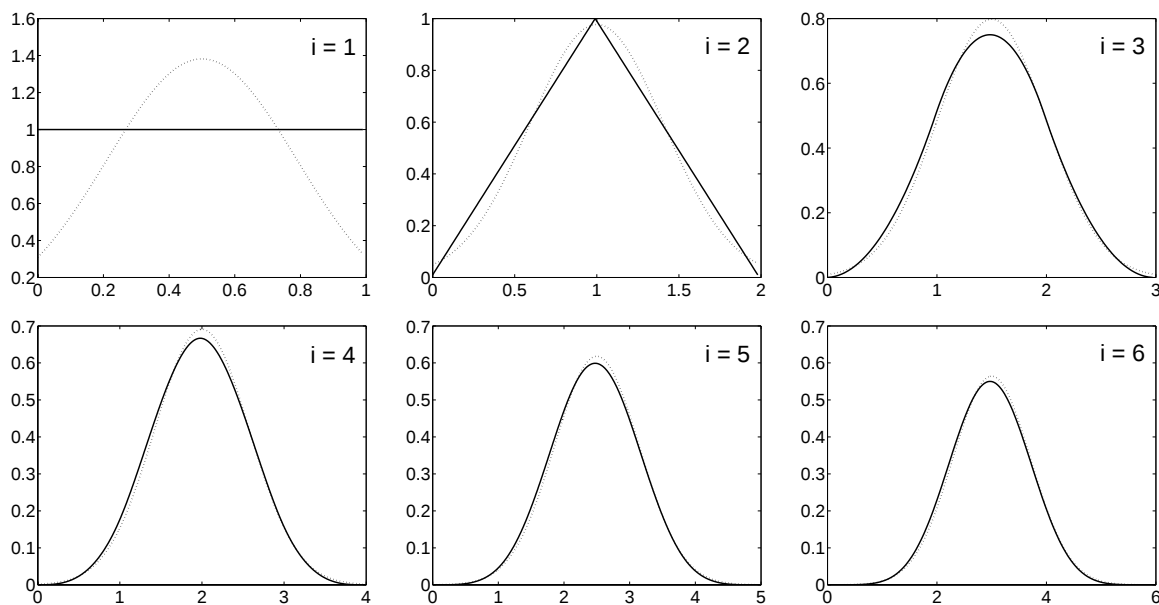


Figura 4.5: Distribuțiile variabilelor aleatoare obținute prin convolvarea a i distribuții uniforme în $[0, 1]$ ($i = 1, \dots, 6$) — reprezentate cu linie continuă — și distribuțiile gaussiene având aceleași medie și dispersie — cu linie punctată.

Nu vom da o demonstrație a acestei teoreme, ci ne vom mărgini doar la argumenta că distribuția lui η , conform relației (4.29), se obține prin N operații de convoluție a distribuțiilor variabilelor ξ_i .

Să considerăm, spre exemplu, un caz particular, și anume să facem ipoteza că ξ_i au toate o distribuție uniformă în intervalul $[0, 1]$. În figura 4.5 sunt prezentate funcțiile rezultate prin convolvarea succesivă a $i = 2, 3 \dots$ distribuții uniforme. Se observă că pe măsură ce i crește, distribuția rezultată variază din ce în ce mai puțin de la i la $i + 1$, tinzând la o funcție invariantă la convoluție. *Această funcție este funcția clopot a lui Gauss.* Se poate demonstra cu ușurință că prin operația de convoluție între două funcții clopot ale lui Gauss (3.23) se obține tot o funcție clopot a lui Gauss, ceea ce justifică rezultatul enunțat de teorema limită centrală.

Nu putem încheia discuția fără a mai face două observații pe tema teoremei limită centrală:

- din cele discutate mai sus, rezultă că suma *oricâtor* variabile aleatoare gaussiene este fără nici un fel de aproximare tot o variabilă aleatoare gaussiană;
- deși teoretic distribuția sumei se apropie de gaussiană doar pentru $N \rightarrow \infty$, se observă că, practic, distribuția rezultată prin sumarea a N variabile aleatoare independente este aproximată destul de bine gaussiană chiar și pentru N mic: se observă din figura 4.5 că diferența corespunzătoare lui $N = 6$ este absolut nesemnificativă.

4.7 Teorema de medie

La fel ca și în cazul unidimensional, se poate enunța și aici o teoremă de medie (pe care-o vom accepta fără demonstrație), care permite calculul mediei variabilei aleatoare transformate ζ direct din distribuția de ordinul doi a perechii inițiale $w_{\xi\eta}$ și din funcția de trecere g .

Teoremă. Media variabilei aleatoare $\zeta = g(\xi, \eta)$, indiferent de funcția g și de distribuția $w_{\xi\eta}$, se poate calcula cu formula:

$$\bar{\zeta} = \iint_{\mathbb{R}^2} g(x, y) w_{\xi\eta}(x, y) dx dy. \quad (4.30)$$

Exemplu. Media sumei a două variabile aleatoare. Fie $\zeta = \xi + \eta$. Avem:

$$\begin{aligned} \bar{\zeta} &= \overline{\xi + \eta} = \iint_{\mathbb{R}^2} (x + y) w_{\xi\eta}(x, y) dx dy \\ &= \iint_{\mathbb{R}^2} x w_{\xi\eta}(x, y) dx dy + \iint_{\mathbb{R}^2} y w_{\xi\eta}(x, y) dx dy \\ &= \int_{-\infty}^{\infty} x \underbrace{\left(\int_{-\infty}^{\infty} w_{\xi\eta}(x, y) dy \right)}_{w_{\xi}(x)} dx + \int_{-\infty}^{\infty} y \underbrace{\left(\int_{-\infty}^{\infty} w_{\xi\eta}(x, y) dx \right)}_{w_{\eta}(y)} dy \\ &= \underbrace{\int_{-\infty}^{\infty} x w_{\xi}(x) dx}_{\bar{\xi}} + \underbrace{\int_{-\infty}^{\infty} y w_{\eta}(y) dy}_{\bar{\eta}} = \bar{\xi} + \bar{\eta} \end{aligned} \quad (4.31)$$

Rezultatul obținut mai sus, conform căruia media sumei a două variabile aleatoare este egală cu suma mediilor, indiferent de gradul de (in)dependență al acestora, este un rezultat important, pe care îl vom folosi intensiv în restul lucrării.

4.8 Momente

După modelul de la capitolul 3, se pot defini momente care să caracterizeze o pereche de variabile aleatoare. Astfel, se definesc momentele necentrate de ordin (k, l) ale perechii de variabile aleatoare (ξ, η) ca fiind media statistică a produsului $\xi^k \eta^l$:

$$m_{\xi\eta}^{(k,l)} \triangleq \overline{\xi^k \eta^l} = \iint_{\mathbb{R}^2} x^k y^l w_{\xi\eta}(x, y) dx dy. \quad (4.32)$$

Momentul necentrat mixt de ordin doi (corespunzător lui $k = l = 1$) se numește *corelația* între ξ și η , și se notează cu $R_{\xi\eta}$:

$$m_{\xi\eta}^{(1,1)} = \overline{\xi\eta} \stackrel{\text{not}}{=} R_{\xi\eta} = \iint_{\mathbb{R}^2} xyw_{\xi\eta}(x,y)dxdy. \quad (4.33)$$

De asemenea, se pot defini momentele centrate de ordin (k, l) ale perechii (ξ, η) ca fiind momentele necentrate ale variabilelor centrate (adică ale variabilelor din care s-a scăzut media, astfel încât să devină de medie nulă):

$$M_{\xi\eta}^{(k,l)} \triangleq \overline{(\xi - \bar{\xi})^k (\eta - \bar{\eta})^l} = \iint_{\mathbb{R}^2} (x - \bar{\xi})^k (y - \bar{\eta})^l w_{\xi\eta}(x,y)dxdy. \quad (4.34)$$

Momentul centrat mixt de ordin doi (pentru $k = l = 1$) se numește *covariația* între ξ și η , și se notează cu $K_{\xi\eta}$:

$$\overline{(\xi - \bar{\xi})(\eta - \bar{\eta})} \stackrel{\text{not}}{=} K_{\xi\eta} = \iint_{\mathbb{R}^2} (x - \bar{\xi})(y - \bar{\eta})w_{\xi\eta}(x,y)dxdy. \quad (4.35)$$

Prin prelucrări elementare, se poate arăta că:

$$\begin{aligned} K_{\xi\eta} &= \overline{(\xi - \bar{\xi})(\eta - \bar{\eta})} = \overline{\xi\eta - \xi\bar{\eta} - \bar{\xi}\eta + \bar{\xi}\bar{\eta}} \\ &= \bar{\xi}\bar{\eta} - \bar{\xi}\bar{\eta} - \bar{\xi}\bar{\eta} + \bar{\xi}\bar{\eta} = \bar{\xi}\bar{\eta} - \bar{\xi}\bar{\eta} - \bar{\xi}\bar{\eta} + \bar{\xi}\bar{\eta} = \bar{\xi}\bar{\eta} - \bar{\xi}\bar{\eta} = R_{\xi\eta} - \bar{\xi}\bar{\eta}. \end{aligned} \quad (4.36)$$

În calculul de mai sus s-au folosit două rezultate obținute anterior: media sumei este suma mediilor, respectiv constantele ies în afara operației de mediere (motiv pentru care, de exemplu, $\overline{\xi\bar{\eta}} = \bar{\xi}\bar{\eta}$).

Definiție. Două variabile aleatoare se numesc *decorelate* atunci când covariația dintre ele este nulă:

$$K_{\xi\eta} = 0. \quad (4.37)$$

Proprietate. Dacă două variabile aleatoare sunt independente, atunci ele sunt și decorelate. Reciproca nu este în general valabilă.

Demonstrație. Dacă ξ și η sunt independente, avem:

$$\begin{aligned} \bar{\xi}\bar{\eta} &= \iint_{\mathbb{R}^2} xyw_{\xi\eta}(x,y)dxdy = \iint_{\mathbb{R}^2} xyw_{\xi}(x)w_{\eta}(y)dxdy \\ &= \underbrace{\int_{-\infty}^{\infty} xw_{\xi}(x)dx}_{\bar{\xi}} \underbrace{\int_{-\infty}^{\infty} yw_{\eta}(y)dy}_{\bar{\eta}} = \bar{\xi}\bar{\eta}, \end{aligned} \quad (4.38)$$

și, deci:

$$K_{\xi\eta} = \bar{\xi}\bar{\eta} - \bar{\xi}\bar{\eta} = 0.$$

Vom arăta în paragraful următor că gradul de corelare între două variabile aleatoare (măsură dată atât de corelația, cât și de covariația între cele două variabile) reprezintă gradul de *dependență liniară* între cele două variabile, adică măsura în care dependența dintre cele două variabile poate fi aproximată printr-o relație liniară de tip $\eta = a\xi + b$.

4.9 Dreapta de regresie

În acest paragraf vom arăta faptul că momentele mixte de ordinul doi ale unei perechi de variabile aleatoare (corelația, covariația) sunt o măsură a gradului de dependență liniară între cele două variabile.

Să presupunem următorul exemplu: ne interesează studiul dependenței statistice între două variabile aleatoare. În absența cunoștințelor statistice despre ξ și η , se acumulează prin experimentări succesive cât mai multe realizări particulare $(\xi^{(i)}, \eta^{(i)})$ ale perechii de variabile aleatoare, iar punctele rezultate se așează pe un grafic bidimensional.

Din forma norului de puncte rezultat se pot trage concluzii referitoare la tipul de dependență între ξ și η . Astfel, cu cât forma norului seamănă mai pronunțat cu o dreaptă (ca în figura 4.6.(a)) cu atât înseamnă că între cele două variabile există o dependență liniară mai puternică, cu alte cuvinte, aproximarea uneia din cealaltă printr-o relație liniară este mai bună. Dacă forma norului este haotică (ca în figura 4.6.(b)), înseamnă că ξ și η sunt independente, și încercarea de a o aproxima pe vreuna din cealaltă după o relație oarecare (nu neapărat liniară) este sortită eșecului. Pot exista cazuri în care norul de puncte nu este haotic, dar forma după care se orientează nu este liniară (v. figura 4.6.(c)); în acest caz, variabilele aleatoare sunt independente liniar (ceea ce e echivalent cu faptul că ele sunt decorelate, după cum vom arăta în continuare) dar nu sunt independente!

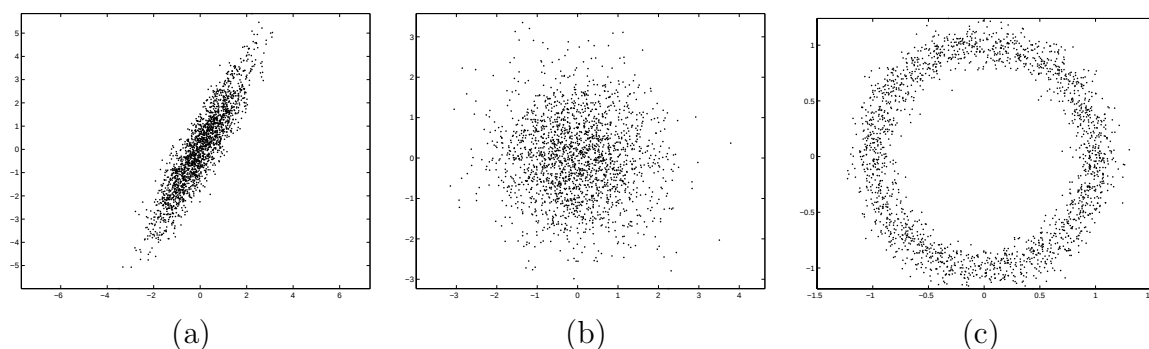


Figura 4.6: Exemplificarea tipurilor de dependență între variabile aleatoare. Diagrama (ξ, η) pentru diverse cazuri (a) ξ și η corelate; (b) ξ și η independente; (c) ξ și η decorelate, dar dependente.

Să presupunem în continuare că ne aflăm în situația în care forma norului de puncte $(\xi^{(i)}, \eta^{(i)})$ este cât de cât liniară; se poate pune, deci, problema aproximării acestuia cât mai bine cu o dreaptă. Dreapta respectivă (numită *dreaptă de regresie*) ar aproxima pe η cu o funcție liniară de ξ (vezi figura 4.7).

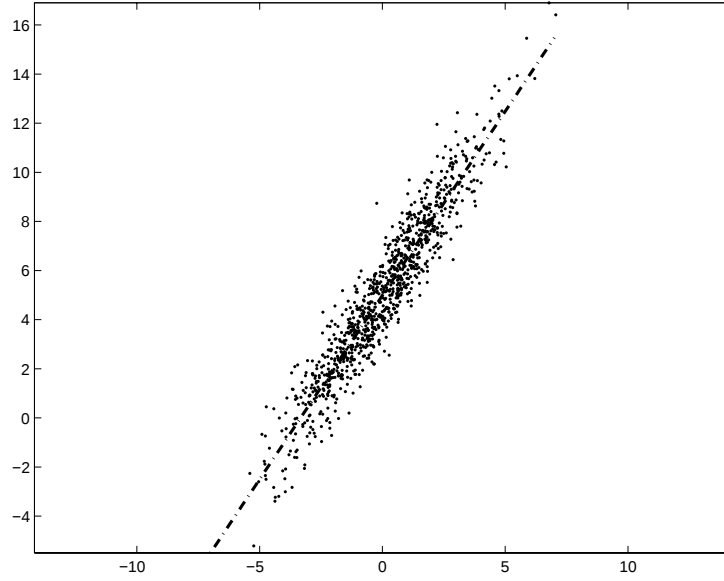


Figura 4.7: Problema aproximării liniare: norul de puncte $(\xi^{(i)}, \eta^{(i)})$ și, cu linie punctată, dreapta de regresie.

Astfel, trebuie căutate numerele $a, b \in \mathbb{R}$ astfel încât estimarea liniară a lui η din ξ , notată cu $\hat{\eta}$ și dată de:

$$\hat{\eta} = a\xi + b \quad (4.39)$$

să se fie “cât mai apropiată posibil” de valoarea reală a lui η . Acest deziderat trebuie tradus matematic, printr-un criteriu de eroare, pe care-l alegem criteriul *erorii pătratice medii minime*. Această alegere este justificată în principal prin faptul că acest criteriu conduce la o soluție matematică ce poate fi exprimată analitic. Se definește, deci, eroarea pătratică medie ε ca fiind:

$$\varepsilon = \overline{(\eta - \hat{\eta})^2} \quad (4.40)$$

cu ajutorul căreia definim matematic problema ca:

$$a, b = ? \text{ astfel încât } \varepsilon = \varepsilon_{\min} \quad (4.41)$$

Problema este una de aflare a minimului unei funcții în funcție de două argumente. Având în vedere faptul că funcția e pătratică în oricare din cele două argumente, se poate arăta că problema se poate trata separat, ca două probleme de minimizare unidimensională. Așadar, valorile lui a și b care-l minimizează pe ε se obțin făcând:

$$\frac{\partial \varepsilon}{\partial b} = 0 \Rightarrow b_o \quad (4.42a)$$

$$\frac{\partial \varepsilon}{\partial a} = 0 \Rightarrow a_o \quad (4.42b)$$

unde a_o și b_o reprezintă valorile optime ale parametrilor a și b . Având în vedere că ε poate fi scrisă ca

$$\varepsilon = \overline{(\eta - \hat{\eta})^2} = \overline{(\eta - a\xi - b)^2}, \quad (4.43)$$

relația (4.42a) devine

$$\frac{\partial \varepsilon}{\partial b} = \frac{\overline{(\eta - a\xi - b)^2}}{\partial b} = -2\overline{(\eta - a\xi - b)} = 0, \quad (4.44)$$

de unde rezultă valoarea lui b :

$$b_o = \bar{\eta} - a\bar{\xi} \quad (4.45)$$

Înlocuind-o pe aceasta în (4.43), rezultă:

$$\varepsilon = \overline{((\eta - \bar{\eta}) - a(\xi - \bar{\xi}))^2}, \quad (4.46)$$

după care (4.42b) poate fi scrisă

$$\begin{aligned} \frac{\partial \varepsilon}{\partial a} &= \frac{\overline{((\eta - \bar{\eta}) - a(\xi - \bar{\xi}))^2}}{\partial a} = -2\overline{((\eta - \bar{\eta}) - a(\xi - \bar{\xi}))(\xi - \bar{\xi})} \\ &= -2\left(\overline{(\eta - \bar{\eta})(\xi - \bar{\xi})} - a\overline{(\xi - \bar{\xi})^2}\right) = -2(K_{\xi\eta} - a\sigma_\xi^2) = 0, \end{aligned} \quad (4.47)$$

de unde rezultă valoarea lui a care-l minimizează pe ε :

$$a_o = \frac{K_{\xi\eta}}{\sigma_\xi^2}. \quad (4.48)$$

Înlocuind valorile obținute pentru a_o și b_o (din relațiile (4.48) și (4.45)) în relația (4.39), rezultă că ecuația dreptei de regresie, care aproximează optim pe η ca o funcție liniară de ξ (în sensul erorii pătratice medii minime) este:

$$\hat{\eta}_o = \frac{K_{\xi\eta}}{\sigma_\xi^2}(\xi - \bar{\xi}) + \bar{\eta}. \quad (4.49)$$

Am notat valoarea optimă a lui $\hat{\eta}$ cu $\hat{\eta}_o$.

Prima observație importantă pe care o facem este că dacă ξ și η sunt decorelate, adică dacă $K_{\xi\eta} = 0$, atunci avem $\hat{\eta}_o = \bar{\eta}$, adică *cea mai bună aproximare liniară a lui η în funcție de ξ nu depinde de acesta*. Altfel spus, între două variabile decorelate nu există nici un fel de dependență liniară!

A doua observație, nu mai puțin importantă, este următoarea: dacă $K_{\xi\eta} \neq 0$ atunci singurul lucru ce poate fi afirmat este că ξ și η nu sunt decorelate, între ele există o anumită dependență liniară. Ce *nu* se poate evalua, însă, pe baza valorii efective a lui $K_{\xi\eta}$ este cât de puternică este dependența respectivă, cu alte cuvinte, cât de apropiat este norul de puncte $(\xi^{(i)}, \eta^{(i)})$ de o dreaptă. Această măsură poate fi studiată numai studiind valoarea erorii obținute prin aproximarea lui η cu $\hat{\eta}_o$.

4.9.1 Coeficientul de corelație

Scriind formula erorii pătratice medii minime obținute, avem:

$$\begin{aligned}
 \varepsilon_{\min} &= \overline{(\eta - \hat{\eta}_o)^2} = \overline{\left((\eta - \bar{\eta}) - \frac{K_{\xi\eta}}{\sigma_\xi^2} (\xi - \bar{\xi}) \right)^2} \\
 &= \underbrace{\overline{(\eta - \bar{\eta})^2}}_{\sigma_\eta^2} - 2 \frac{K_{\xi\eta}}{\sigma_\xi^2} \underbrace{\overline{(\eta - \bar{\eta})(\xi - \bar{\xi})}}_{K_{\xi\eta}} + \frac{K_{\xi\eta}^2}{\sigma_\xi^4} \underbrace{\overline{(\xi - \bar{\xi})^2}}_{\sigma_\xi^2} \\
 &= \sigma_\eta^2 - 2 \frac{K_{\xi\eta}^2}{\sigma_\xi^2} + \frac{K_{\xi\eta}^2}{\sigma_\xi^2} = \sigma_\eta^2 - \frac{K_{\xi\eta}^2}{\sigma_\xi^2} \\
 &= \sigma_\eta^2 \left[1 - \left(\frac{K_{\xi\eta}}{\sigma_\xi \sigma_\eta} \right)^2 \right].
 \end{aligned} \tag{4.50}$$

Termenul pus în evidență în expresia finală a erorii pătratice medii minime se notează cu $\rho_{\xi\eta}$ și se numește *coeficient de corelație*:

$$\rho_{\xi\eta} = \frac{K_{\xi\eta}}{\sigma_\xi \sigma_\eta}. \tag{4.51}$$

Coeficientul de corelație $\rho_{\xi\eta}$ reprezintă valoarea normalizată a covariației, care permite evaluarea absolută a măsurii dependenței liniare între ξ și η . Într-adevăr, scriind:

$$\varepsilon_{\min} = \sigma_\eta^2 (1 - \rho_{\xi\eta}^2), \tag{4.52}$$

se observă că $\rho_{\xi\eta}$ este de modul subunitar, în caz contrar, eroarea pătratică medie devenind negativă, ceea ce este imposibil, având în vedere că aceasta reprezintă media pătratului unei valori:

$$|\rho_{\xi\eta}| \leq 1. \tag{4.53}$$

Mai mult, se observă că pe măsură ce $|\rho_{\xi\eta}| \nearrow 1$, eroarea pe care o facem aproximându-l pe η cu $\hat{\eta}_o$ este din ce în ce mai mică. La limită, avem:

$$|\rho_{\xi\eta}| = 1 \quad \Leftrightarrow \quad \varepsilon_{\min} = 0 \quad \Leftrightarrow \quad \eta = \hat{\eta}_o,$$

cu alte cuvinte, dacă modulul coeficientului de corelație este egal cu unu, între ξ și η există o dependență liniară funcțională (norul de puncte $(\xi^{(i)}, \eta^{(i)})$ este o dreaptă)¹.

O ultimă observație la acest capitol este legată de semnificația unei valori negative a coeficientului de corelație. Faptul că $\rho_{\xi\eta} < 0$ semnifică numai faptul că panta dreptei de regresie este negativă, în nici un caz faptul că ξ și η sunt mai puțin corelate decât în cazul $\rho_{\xi\eta} \geq 0$!

În figura 4.8, este prezentat un exemplu al formei norului de puncte $(\xi^{(i)}, \eta^{(i)})$ pentru diverse valori ale lui $\rho_{\xi\eta}$.

¹Dacă media pătratică a unei variabile aleatoare este nulă, atunci variabila aleatoare respectivă (în cazul nostru $\eta - \eta_o$) este determinist egală cu 0.

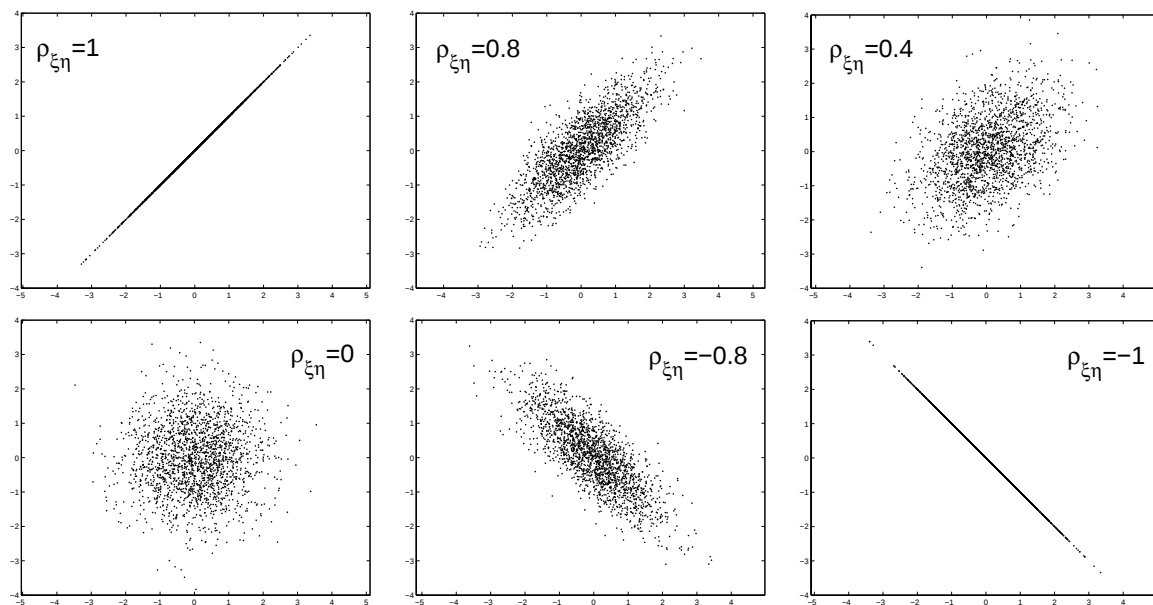


Figura 4.8: Forma norului de puncte $(\xi^{(i)}, \eta^{(i)})$ pentru diverse valori ale coeficientului de corelație.

4.10 Distribuția gaussiană de ordinul doi

Variabilele aleatoare ξ și η se zice că au o distribuție gaussiană (sau normală) de ordinul doi de parametri $m_1, m_2, \sigma_1, \sigma_2, r$ (și se notează $(\xi, \eta): \mathcal{N}(m_1, m_2, \sigma_1, \sigma_2, r)$) dacă densitatea lor de probabilitate de ordinul doi este dată de:

$$w_{\xi\eta}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp \left[-\frac{1}{2(1-r^2)} \left(\frac{(x-m_1)^2}{\sigma_1^2} - 2r \frac{(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \frac{(y-m_2)^2}{\sigma_2^2} \right) \right], \quad (4.54)$$

cu m_1 și m_2 oarecare, $\sigma_1 > 0$, $\sigma_2 > 0$ și $|r| \leq 1$.

Prin calcul direct, folosind proprietatea 5 a densității de probabilitate de ordinul doi, se poate arăta că dacă perechea (ξ, η) are o distribuție gaussiană de ordinul doi, atunci și ξ și η au o distribuție gaussiană de ordinul unu, mai precis $\xi: \mathcal{N}(m_1, \sigma_1)$, respectiv $\eta: \mathcal{N}(m_2, \sigma_2)$.

Reciproca afirmației de mai sus nu e valabilă. Cu alte cuvinte, faptul că atât ξ cât și η au fiecare o distribuție gaussiană de ordinul unu nu garantează faptul că perechea (ξ, η) are o distribuție gaussiană de ordinul doi!

Am văzut mai sus că primii patru parametri ai gaussienei de ordinul doi au semnificația de medii, respectiv dispersii, ale celor două variabile aleatoare. Cât despre al cincilea parametru, respectiv r , se arată prin calcul direct că acesta reprezintă coeficientul de

corelație între ξ și η :

$$\rho_{\xi\eta} = \frac{K_{\xi\eta}}{\sigma_\xi\sigma_\eta} = \dots = r. \quad (4.55)$$

Deci, forma clopotului lui Gauss bidimensional va fi mai mult sau mai puțin elongată în funcție de cât de apropiat este $|r|$ de 1 (respectiv de 0). În figura 4.9 se arată forma funcției (4.54) pentru două valori ale lui r .

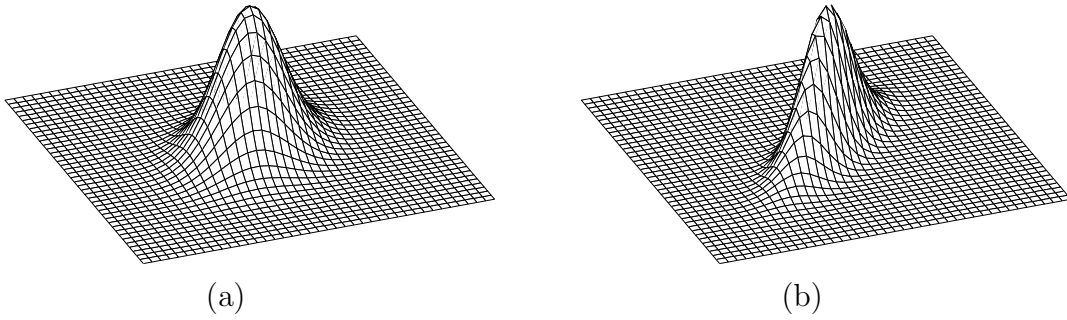


Figura 4.9: Forma distribuției gaussiene de ordinul doi pentru: (a) $r = 0,3$ și (b) $r = 0,8$.

Din faptul că $\rho_{\xi\eta} = r$ mai decurge o altă proprietate importantă a gaussienei. Pentru $r = 0$, se poate scrie:

$$\begin{aligned} w_{\xi\eta}(x, y) &= \frac{1}{2\pi\sigma_1\sigma_2} \exp \left[-\frac{1}{2} \left(\frac{(x - m_1)^2}{\sigma_1^2} + \frac{(y - m_2)^2}{\sigma_2^2} \right) \right] \\ &= \frac{1}{\sigma_1\sqrt{2\pi}} \exp \left(-\frac{(x - m_1)^2}{2\sigma_1^2} \right) \frac{1}{\sigma_2\sqrt{2\pi}} \exp \left(-\frac{(y - m_2)^2}{2\sigma_2^2} \right) \\ &= w_\xi(x)w_\eta(y). \end{aligned} \quad (4.56)$$

Altfel spus, dacă două variabile aleatoare cu distribuție gaussiană de ordinul doi sunt decorelate, atunci ele sunt și independente!

Capitolul 5

Semnale aleatoare

5.1 Introducere

După ce am studiat caracterizarea statistică a variabilelor aleatoare și a perechilor de variabile aleatoare, este momentul să mergem mai departe. Vom aborda în continuare problema semnalelor aleatoare, care pot fi văzute precum “colecții” de variabile aleatoare cu desfășurare în timp. Practic, vom vedea cum se aplică toate rezultatele obținute în capitolele anterioare pentru caracterizarea statistică a semnalelor aleatoare, care sunt obiectul principal de studiu al acestei lucrări.

Semnalele aleatoare sunt semnale guvernate, fie și parțial, de legi probabilistice. Cu alte cuvinte, valoarea unui semnal aleator la un moment dat este, cel puțin parțial, *impredictibilă*. Din punct de vedere matematic, un semnal aleator este o funcție definită atât pe spațiul eșantioanelor Ω cât și pe $\mathbb{R}$:

$$\xi : \Omega \times \mathbb{R} \longrightarrow \mathbb{R}. \quad (5.1)$$

Dacă fixăm $\omega_k \in \Omega$, atunci avem de-a face cu o *realizare particulară* a semnalului aleator, $\xi(\omega_k, t) \stackrel{\text{not}}{=} \xi^{(k)}(t)$, care este ceea ce convențional numim un semnal, adică o funcție de timp. Este momentul să atragem atenția, pentru a evita confuziile și a fixa noțiunile, asupra faptului că termenul de semnal aleator se referă la o *clasă de semnale* caracterizate de aceleași proprietăți statistice, și nu la vreun semnal în parte, acesta din urmă numindu-se, așa cum am precizat mai sus, realizare particulară a semnalului aleator. Aceasta este ilustrată în figura 5.1.

5.2 Caracterizarea statistică a semnalelor aleatoare

Dacă fixăm un moment de timp oarecare t_1 , valoarea semnalului la momentul respectiv de timp este o *variabilă aleatoare* $\xi(t_1) : \Omega \longrightarrow \mathbb{R}$. Această observație ne înlesnește descrierea statistică a semnalului aleator folosind mărimile statistice introduse în capitolele anterioare pentru caracterizarea variabilelor aleatoare sau a perechilor de variabile aleatoare.

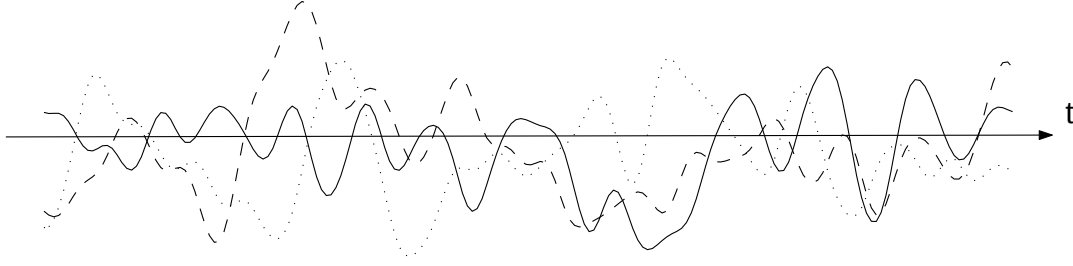


Figura 5.1: Diverse realizări particulare ale unui semnal aleator.

5.2.1 Caracterizarea de ordinul unu a semnalelor aleatoare

Se definește funcția de repartiție de ordinul unu a semnalului aleator $\xi(t)$ ca fiind:

$$F_{\xi}(x_1, t_1) \triangleq P(\xi(t_1) \leq x_1) \quad (5.2)$$

Se observă că definiția (5.2) este similară celei din (3.3), cu diferența că, în cazul semnalelor, funcția de repartiție depinde și de timp. Acest lucru este normal, întrucât funcția de repartiție trebuie să caracterizeze întreg semnalul, adică variabilele aleatoare $\xi(t_1)$ pentru *orice* moment de timp $t_1 \in \mathbb{R}$.

Apoi, se definește densitatea de probabilitate de ordinul unu a semnalului, după:

$$w_{\xi}(x_1, t_1) = \frac{\partial F_{\xi}(x_1, t_1)}{\partial x_1}, \quad (5.3)$$

pe baza căreia se pot calcula momentele statistice ale semnalului, printre care și media, media pătratică și varianța acestuia:

$$\overline{\xi(t_1)} = \int_{-\infty}^{\infty} x_1 w_{\xi}(x_1, t_1) dx_1, \quad (5.4)$$

$$\overline{\xi^2(t_1)} = \int_{-\infty}^{\infty} x_1^2 w_{\xi}(x_1, t_1) dx_1, \quad (5.5)$$

$$\sigma_{\xi}^2(t_1) = \overline{\xi^2(t_1)} - \overline{\xi(t_1)}^2. \quad (5.6)$$

Toate aceste mărimi statistice de ordinul unu permit caracterizarea completă a comportamentului statistic al semnalului la orice moment de timp. Ceea ce nu poate fi evaluat, însă, cu ajutorul acestor mărimi este interdependența statistică între valorile semnalului la două momente de timp diferite. Pentru aceasta este nevoie de mărimi statistice de ordinul doi, pe care le vom discuta în paragraful următor.

5.2.2 Caracterizarea de ordinul doi a semnalelor aleatoare

Se definește funcția de repartiție de ordinul doi a semnalului aleator $\xi(t)$ ca fiind funcția de repartiție de ordinul doi a variabilelor aleatoare $\xi(t_1)$ și $\xi(t_2)$:

$$F_{\xi\xi}(x_1, x_2, t_1, t_2) \triangleq P((\xi(t_1) \leq x_1) \cap (\xi(t_2) \leq x_2)). \quad (5.7)$$

Ca și pentru ordinul unu, funcția de repartiție de ordinul doi depinde și de momentele de timp t_1 și t_2 . Se poate defini apoi densitatea de probabilitate de ordinul doi a semnalului:

$$w_{\xi\xi}(x_1, x_2, t_1, t_2) = \frac{\partial^2 F_{\xi\xi}(x_1, x_2, t_1, t_2)}{\partial x_1 \partial x_2}, \quad (5.8)$$

care permite calculul momentelor mixte ale acestuia. Se definește, deci, funcția de autocorelație a semnalului ca fiind corelația dintre $\xi(t_1)$ și $\xi(t_2)$, $\forall t_1, t_2 \in \mathbb{R}$:

$$R_\xi(t_1, t_2) \triangleq \overline{\xi(t_1)\xi(t_2)} = \iint_{\mathbb{R}^2} x_1 x_2 w_{\xi\xi}(x_1, x_2, t_1, t_2) dx_1 dx_2. \quad (5.9)$$

Facem observația că, deși în spiritul notațiilor utilizate până în acest punct, funcția de autocorelație ar trebui notată cu $R_{\xi\xi}$, noi o vom nota cu R_ξ , considerând că această simplificare a notației nu introduce nici un echivoc.

În mod absolut similar, putem defini funcția de autocovariație a semnalului ca fiind funcția de autocorelație a valorilor centrate ale semnalului:

$$K_\xi(t_1, t_2) \triangleq \overline{(\xi(t_1) - \overline{\xi(t_1)}) (\xi(t_2) - \overline{\xi(t_2)})} = \overline{\xi(t_1)\xi(t_2)} - \overline{\xi(t_1)} \overline{\xi(t_2)}. \quad (5.10)$$

Descrierea statistică a semnalelor aleatoare ar putea continua prin definirea densităților de probabilitate de ordin superior lui doi. Evident, cu cât ordinul este mai mare, cu atât descrierea semnalului este mai completă. Pe de altă parte, cu cât ordinul este mai mare, cu atât devin mai greu de estimat în mod practic mărimile respective. Din aceste motive, majoritatea aplicațiilor în prelucrarea statistică a semnalelor folosesc mărimile de ordin cel mult doi (medie, funcție de autocorelație, etc.). Există, totuși, și aplicații mai speciale ale prelucrării semnalelor, în care este necesară folosirea mărimilor statistice de ordin trei și patru, numite *statistici de ordin superior*, pentru a obține rezultate ce nu pot fi extrase folosind mărimile de ordinul unu și doi.

5.3 Semnale staționare

În acest paragraf, vom introduce o noțiune foarte importantă în teoria statistică a semnalelor, și anume staționaritatea semnalelor aleatoare. Există două feluri de staționaritate, în sens strict și în sens larg, după cum vom arăta în paragrafele următoare.

5.3.1 Staționaritate în sens strict

Un semnal aleator $\xi(t)$ se zice că este staționar în sens strict de ordinul n dacă densitatea sa de probabilitate de până la ordinul n este invariantă la o translație oarecare în timp:

$$\underbrace{w_{\xi \dots \xi}}_n(x_1, \dots, x_n, t_1, \dots, t_n) = \underbrace{w_{\xi \dots \xi}}_n(x_1, \dots, x_n, t_1 + \tau, \dots, t_n + \tau), \quad \forall \tau \in \mathbb{R}. \quad (5.11)$$

Evident, relația de mai sus este valabilă pentru orice $n' < n$, întrucât densitatea de ordin n' se obține ca densitate marginală a celei de ordin n .

În particular, un semnal staționar în sens strict de ordinul doi este un semnal pentru care:

$$w_{\xi}(x_1, t_1) = w_{\xi}(x_1, t_1 + \tau), \quad \forall \tau \in \mathbb{R} \quad (5.12)$$

$$w_{\xi\xi}(x_1, x_2, t_1, t_2) = w_{\xi\xi}(x_1, x_2, t_1 + \tau, t_2 + \tau), \quad \forall \tau \in \mathbb{R} \quad (5.13)$$

Dacă alegem $\tau = -t_1$ în relația (5.12), obținem:

$$w_{\xi}(x_1, t_1) = w_{\xi}(x_1, 0) = w_{\xi}(x_1), \quad (5.14)$$

cu alte cuvinte, prima consecință directă a staționarității unui semnal este faptul că densitatea sa de probabilitate de ordinul unu *nu depinde de timp*: la orice moment de timp t , variabila aleatoare $\xi(t)$ are *aceeași* densitate de probabilitate. Consecința imediată a acestui fapt este și faptul că media, media pătratică și dispersia semnalului au aceeași proprietate. Într-adevăr, dacă înlocuim (5.14) în relațiile, (5.4)–(5.6), obținem:

$$\overline{\xi(t_1)} = \bar{\xi}, \quad (5.15)$$

$$\overline{\xi^2(t_1)} = \bar{\xi^2}, \quad (5.16)$$

$$\sigma_{\xi}^2(t_1) = \sigma_{\xi}^2. \quad (5.17)$$

Dacă, în plus, alegem $\tau = -t_2$ în ecuația (5.13), obținem:

$$w_{\xi\xi}(x_1, x_2, t_1, t_2) = w_{\xi\xi}(x_1, x_2, t_1 - t_2, 0) = w_{\xi\xi}(x_1, x_2, t_1 - t_2), \quad \forall \tau \in \mathbb{R}. \quad (5.18)$$

Relația (5.18) ne arată că densitatea de probabilitate de ordinul doi a semnalului staționar în sens strict de ordinul doi nu depinde de cele două momente de timp ca atare, ci numai de diferența între ele. Altfel spus, oricare două valori ale semnalului localizate la același ecart temporal una față de cealaltă vor avea aceeași distribuție de ordinul doi, indiferent de localizarea lor absolută pe axa timpului. Evident, acest lucru se va răsfrânge asupra momentelor mixte ale semnalului, adică funcțiile de autocorelație și autocovariație, care vor avea și ele aceeași proprietate de a nu depinde de cele două argumente ca atare, ci numai de diferența între ele. Astfel, introducând (5.18) în relațiile (5.9) și (5.10), obținem:

$$R_{\xi}(t_1, t_2) = R_{\xi}(t_1 - t_2), \quad (5.19)$$

$$K_{\xi}(t_1, t_2) = K_{\xi}(t_1 - t_2), \quad (5.20)$$

cu

$$R_{\xi}(\tau) = \overline{\xi(t)\xi(t+\tau)} \quad \forall t \in \mathbb{R}, \quad (5.21)$$

$$K_{\xi}(\tau) = R_{\xi}(\tau) - \bar{\xi}^2. \quad (5.22)$$

Se observă din relația (5.22) că în ipoteza staționarității, funcțiile de autocorelație și autocovariație diferă numai printr-o constantă. Dacă, în plus, semnalul este de medie nulă, atunci cele două funcții coincid.

5.3.2 Staționaritate în sens larg

Staționaritatea în sens strict este o ipoteză greu de verificat în practică pentru majoritatea semnalelor reale, chiar și pentru valori mici ale lui n . În practică, se folosește o ipoteză simplificată, și anume cea de staționaritate în sens larg, care implică invarianța la o translație în timp nu a densităților de probabilitate, ci a momentelor semnalului de ordinul unu și doi.

Un semnal se numește staționar în sens larg dacă:

1. media lui este independentă de timp (relația (5.15));
2. funcția lui de autocorelație depinde doar de diferența între cele două argumente ale sale (relația (5.19)).

Se poate arăta că din (5.15) și (5.19) se pot deduce și relațiile de invarianță la o translație în timp a celorlalte momente de interes, cu alte cuvinte că și relațiile (5.16), (5.17) și (5.20) sunt adevărate în ipoteza staționarității în sens larg.

În cele ce urmează, ne vom referi pe scurt la semnalele staționare în sens larg ca fiind *semnale staționare*.

După cum spuneam, clasa semnalelor staționare ocupă un loc foarte important în studiul statistic al semnalelor. Majoritatea semnalelor reale sunt fie staționare, fie pot fi approximate pe durate mai lungi sau mai scurte cu semnale staționare, ceea ce explică interesul arătat pentru acest tip de semnale. În continuarea acestei lucrări, vom considera staționaritatea semnalelor studiate ca fiind o ipoteză de lucru.

5.4 Proprietățile funcției de autocorelație

În acest paragraf, vom stabili câteva proprietăți importante ale funcției de autocorelație a semnalelor aleatoare staționare dată de (5.21).

1. Funcția de autocorelație este pară.

$$R_{\xi}(\tau) = R_{\xi}(-\tau) \quad \forall \tau \in \mathbb{R}. \quad (5.23)$$

Demonstrație. Pornind de la (5.21) și alegând $\tau = t_1 - t$, obținem $R_{\xi}(\tau) = \overline{\xi(t_1 - \tau)\xi(t_1)} = R_{\xi}(-\tau)$.

Este evident că așa și trebuie să fie: $R_{\xi}(\tau)$ reprezintă corelația între două valori ale semnalului decalate în timp cu τ , *oriunde* ar fi ele localizate în mod absolut pe axa timpului. Fie două momente t_1 și t_2 , cu $t_2 - t_1 = \tau$. Atunci, t_2 e decalat cu τ față de t_1 , în timp ce t_1 e decalat cu $-\tau$ față de t_2 . Rezultă că $R_{\xi}(\tau)$ și $R_{\xi}(-\tau)$ măsoară, de fapt, același lucru, și anume corelația dintre variabilele aleatoare $\xi(t_1)$ și $\xi(t_2)$, deci trebuie obligatoriu să fie egale!

2. Funcția de autocorelație este maximă în origine.

$$R_\xi(0) \geq |R_\xi(\tau)| \quad \forall \tau \in \mathbb{R}. \quad (5.24)$$

Demonstrație. Calculăm media pătratică a diferenței $\xi(t) - \xi(t + \tau)$:

$$\begin{aligned} \overline{(\xi(t) - \xi(t + \tau))^2} &= \overline{\xi^2(t) - 2\xi(t)\xi(t + \tau) + \xi^2(t + \tau)} \\ &= \underbrace{\overline{\xi^2(t)}}_{R_\xi(0)} - \underbrace{2\overline{\xi(t)\xi(t + \tau)}}_{R_\xi(\tau)} + \underbrace{\overline{\xi^2(t + \tau)}}_{R_\xi(0)} \\ &= 2(R_\xi(0) - R_\xi(\tau)). \end{aligned} \quad (5.25)$$

Cum media pătratică este obligatoriu pozitivă, rezultă că $R_\xi(0) \geq R_\xi(\tau)$. Repetând raționamentul de mai sus pornind de la $\xi(t) + \xi(t + \tau)$, va rezulta că $R_\xi(0) \geq -R_\xi(\tau)$, ceea ce completează demonstrația relației (5.24).

Relația (5.24) are și ea o interpretare evidentă: valoarea funcției de autocorelație în 0 reprezintă corelația dintre valoarea semnalului la un moment de timp dat și ea însăși. Este evident că această corelație (care e perfectă) nu poate fi mai mică decât corelația respectivei valori cu altă valoare a semnalului, decalată cu un τ oarecare.

3. În ipoteza că în semnal nu există componente periodice sau deterministe, valoarea funcției de autocorelație la infinit este egală cu pătratul mediei semnalului.

$$R_\xi(\infty) = \bar{\xi}^2. \quad (5.26)$$

Demonstrație. Revenind la (5.21) observăm că atunci când $\tau \rightarrow \infty$, avem de-a face cu două valori care, deși provin din același semnal, datorită ecartului temporal între ele, este verosimil să nu mai fie legate statistic una de alta, adică să fie independente. În acest caz, media produsului devine egală cu produsului mediilor, care sunt egale, și deci, relația (5.26) este demonstrată. Evident, presupunerea de independență statistică între două valori ale semnalului foarte îndepărtate temporal una de cealaltă nu mai e valabilă dacă, de exemplu, în semnal există o componentă periodică.

4. Media pătratică și varianța semnalului se obțin din funcția de autocorelație astfel:

$$\bar{\xi}^2 = R_\xi(0) \quad (5.27)$$

$$\sigma_\xi^2 = R_\xi(0) - R_\xi(\infty) \quad (5.28)$$

Demonstrație. Relația (5.27) nu este altceva decât exprimarea relației (5.21) pentru $\tau = 0$, ținând în plus cont de (5.16), în timp ce relația (5.28) decurge firesc din ecuațiile (5.27) și (5.26).

5. Dacă semnalul aleator este periodic, atunci și funcția lui de autocorelație este periodică, având aceeași perioadă ca și semnalul.

$$\xi(t) = \xi(t + T) \quad \forall t \in \mathbb{R} \quad \Rightarrow \quad R_\xi(\tau) = R_\xi(\tau + T) \quad \forall \tau \in \mathbb{R}. \quad (5.29)$$

Demonstrație. Scriem că

$$R_{\xi}(\tau + T) = \overbrace{\xi(t) \xi(t + \tau + T)}^{\xi(t+T)} = \overline{\xi(t+T) \xi(t + \tau + T)} = R_{\xi}(\tau). \quad (5.30)$$

Interpretarea relației (5.29) este de asemenea imediată. Dacă semnalul este periodic, dacă valorile sale se repetă identic după o durată T , este logic să se întâmple același lucru și cu funcția sa de autocorelație: $R_{\xi}(\tau)$ este corelația dintre $\xi(t)$ și $\xi(t + \tau)$, care, datorită faptului că $\xi(t + \tau) = \xi(t + \tau + T)$ este identică cu corelația dintre $\xi(t)$ și $\xi(t + \tau + T)$, care nu este altceva decât $R_{\xi}(\tau + T)$.

5.5 Funcția de intercorelație

Fie $\xi(t)$ și $\eta(t)$ două semnale staționare. Se pot defini funcțiile de intercorelație și intercovariație între cele două semnale prin analogie cu funcțiile similare ce caracterizează un singur semnal. Astfel, pornind de la relația (5.21) se poate scrie funcția de intercorelație între $\xi(t)$ și $\eta(t)$ precum:

$$R_{\xi\eta}(\tau) = \overline{\xi(t)\eta(t+\tau)} \quad \forall t \in \mathbb{R}, \quad (5.31)$$

și, în mod similar, se poate scrie și funcția de intercovariație:

$$K_{\xi\eta}(\tau) = \overline{(\xi(t) - \bar{\xi}(t))(\eta(t+\tau) - \bar{\eta}(t+\tau))} = R_{\xi\eta}(\tau) - \bar{\xi} \bar{\eta}, \quad \forall t \in \mathbb{R} \quad (5.32)$$

Se observă din modul de definire că funcția de intercorelație nu mai este o funcție pară. În schimb, reluând demonstrația de la proprietatea 1 a funcției de autocorelație, se poate arăta că funcția de intercorelație este caracterizată de:

$$R_{\xi\eta}(\tau) = R_{\eta\xi}(-\tau) \quad (5.33)$$

În mod similar cu (5.31), se poate defini și funcția de intercovariație $K_{\xi\eta}(\tau)$.

5.6 Semnale ergodice

În acest paragraf, vom introduce o nouă clasă de semnale, mai restrictivă decât clasa semnalelor staționare, și anume semnalele ergodice. Pentru aceasta, însă, este necesară definirea mediilor temporale ale semnalului.

5.6.1 Medii temporale ale semnalului

Fie $\xi^{(k)}(t)$ o realizare particulară a semnalului aleator staționar $\xi(t)$. Se definește media temporală a realizării particulare ca fiind:

$$\widetilde{\xi^{(k)}} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \xi^{(k)}(t) dt. \quad (5.34)$$

În mod similar, se poate defini o funcție de autocorelație temporală calculată pe realizarea particulară respectivă:

$$R_{\xi^{(k)}}^t(t_1, t_2) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \xi^{(k)}(t + t_1) \xi^{(k)}(t + t_2) dt. \quad (5.35)$$

Făcând schimbarea de variabilă $t + t_2 = \tau$ în ecuația (5.35), se observă că, prin însuși modul ei de definiție, funcția de autocorelație temporală depinde numai de diferența între cele două argumente ale ei:

$$R_{\xi^{(k)}}^t(t_1, t_2) = R_{\xi^{(k)}}^t(t_1 - t_2). \quad (5.36)$$

5.6.2 Ergodicitate în sensul mediei și în sensul autocorelației

Un semnal aleator staționar $\xi(t)$ se numește *ergodic în sensul mediei* dacă media temporală a oricărei realizări particulare ale sale aceeași și este egală cu media statistică:

$$\bar{\xi} = \widetilde{\xi^{(k)}} \quad \forall k. \quad (5.37)$$

Un semnal aleator staționar $\xi(t)$ se numește *ergodic în sensul funcției de autocorelație* dacă funcția sa de autocorelație calculată pe oricare realizare particulară a sa este identică cu funcția de autocorelație statistică:

$$R_{\xi}(\tau) \equiv R_{\xi^{(k)}}^t(\tau) \quad \forall k. \quad (5.38)$$

În primul rând, să facem observația că problema ergodicității nu s-ar putea pune în absența ipotezei de staționaritate a semnalului. (Într-adevăr, dacă semnalul ar fi nestaționar, atunci media lui statistică ar varia cu timpul, deci n-ar fi unică, și, deci, nu s-ar mai putea pune problema egalității cu media temporală a unei realizări particulare oarecare.)

În al doilea rând, este util de precizat faptul că ergodicitatea este o ipoteză făcută de multe ori din lipsă de suficiente date, și nu întotdeauna verificată în practică. Ea permite calculul mărimilor statistice ale semnalului pornind de la o singură realizare particulară a acestuia, și este invocată ori de câte ori avem nevoie de mărimile respective în condițiile în care dispunem de un număr de realizări particulare ale semnalului insuficient pentru o estimare precisă a acestora.

În paragraful următor, vom demonstra o teoremă care ne arată condiția pe care trebuie să o satisfacă funcția de covariație a unui semnal staționar pentru ca el să fie și ergodic în sensul mediei.

5.6.3 Teorema ergodicității mediei

Fie $\xi^{(k)}(t)$ realizarea particulară a semnalului aleator staționar $\xi(t)$ definită mai sus. Fie $\mu_T^{(k)}(\xi)$ media temporală a realizării particulare trunchiate pe intervalul $[-\frac{T}{2}, \frac{T}{2}]$:

$$\mu_T^{(k)}(\xi) = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \xi^{(k)}(t) dt. \quad (5.39)$$

Evident, media temporală definită mai sus este o variabilă aleatoare, întrucât ia în considerare un interval finit pentru integrare, și deci, va varia de la o realizare particulară $\xi^{(k)}$ la alta. Media statistică a lui $\mu_T^{(k)}(\xi)$ este dată de:

$$\overline{\mu_T^{(k)}(\xi)} = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \xi^{(k)}(t) dt = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \underbrace{\xi^{(k)}(t)}_{\bar{\xi}} dt = \frac{1}{T} \bar{\xi} t \Big|_{-\frac{T}{2}}^{\frac{T}{2}} = \bar{\xi}. \quad (5.40)$$

În dezvoltarea de mai sus, am folosit ipoteza de staționaritate a semnalului prin faptul că media statistică este invariantă în timp. De asemenea, am ținut cont de faptul (demonstrat în capitolul 4) că operația de mediere statistică comută cu operația de sumare (integrare); în plus constantele (în cazul de față raportul $\frac{1}{T}$) ies în afara operației de mediere.

Semnalul $\xi(t)$ se este ergodic în sensul mediei dacă:

$$\lim_{T \rightarrow \infty} \overline{\left(\mu_T^{(k)}(\xi) - \bar{\xi} \right)^2} = 0. \quad (5.41)$$

Într-adevăr, relația de mai sus este echivalentă cu definiția (5.37): ea afirmă că varianța variabilei aleatoare $\mu_T^{(k)}(\xi)$ tinde la 0 atunci când $T \rightarrow \infty$. Or, faptul că varianța unei variabile aleatoare este nulă este echivalent cu a spune că variabila respectivă este deterministă, egală cu media ei. Altfel spus, media temporală pe $\mathbb{R}$ a oricărei realizări particulare a semnalului este egală cu media statistică a acestuia.

Dezvoltând termenul din relația (5.41), obținem:

$$\begin{aligned} \overline{\left(\mu_T^{(k)}(\xi) - \bar{\xi} \right)^2} &= \overline{\left(\frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \xi^{(k)}(t) dt - \bar{\xi} \right)^2} = \overline{\left(\frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \xi^{(k)}(t) dt - \frac{1}{T} \bar{\xi} T \right)^2} \\ &= \overline{\left(\frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} (\xi^{(k)}(t) - \bar{\xi}) dt \right)^2} = \overline{\left(\frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} (\xi^{(k)}(t_1) - \bar{\xi}) dt_1 \right) \left(\frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} (\xi^{(k)}(t_2) - \bar{\xi}) dt_2 \right)} \\ &= \frac{1}{T^2} \iint_{-\frac{T}{2}}^{\frac{T}{2}} (\xi^{(k)}(t_1) - \bar{\xi}) (\xi^{(k)}(t_2) - \bar{\xi}) dt_1 dt_2 \\ &= \frac{1}{T^2} \iint_{-\frac{T}{2}}^{\frac{T}{2}} \underbrace{(\xi^{(k)}(t_1) - \bar{\xi}) (\xi^{(k)}(t_2) - \bar{\xi})}_{K_{\xi}(t_1, t_2)} dt_1 dt_2 = \frac{1}{T^2} \iint_{-\frac{T}{2}}^{\frac{T}{2}} K_{\xi}(t_1 - t_2) dt_1 dt_2. \end{aligned} \quad (5.42)$$

În dezvoltarea de mai sus, am identificat funcția de autocovariație a semnalului (5.10), care, datorită staționarității acestuia, depinde numai de diferența dintre argumentele sale (ecuația (5.20)).

În ecuația (5.42) se face următoarea schimbare de variabile:

$$\begin{aligned}\tau &= t_1 - t_2 \\ \theta &= t_2\end{aligned}\tag{5.43}$$

Domeniul pătrat $[-\frac{T}{2}, \frac{T}{2}] \times [-\frac{T}{2}, \frac{T}{2}]$ pe care se face integrarea în coordonatele inițiale (t_1, t_2) devine un paralelogram în noile coordonate (θ, τ) , după cum este prezentat în figura 5.2. Într-adevăr, întrucât $\theta = t_2$, limitele de variație ale lui θ sunt $[-\frac{T}{2}, \frac{T}{2}]$. Pentru un θ fixat, $\tau = t_1 - t_2 = t_1 - \theta$, deci, având în vedere că $t_1 \in [-\frac{T}{2}, \frac{T}{2}]$, rezultă că $\tau \in [-\frac{T}{2} - \theta, \frac{T}{2} - \theta]$, ceea ce explică forma domeniului de integrare din figura 5.2.(b).

Având în vedere că transformarea inversă este:

$$\begin{aligned}t_1 &= \tau + \theta \\ t_2 &= \theta,\end{aligned}\tag{5.44}$$

jacobianul transformării (5.43) este:

$$J(\tau, \theta) = \det \left(\begin{bmatrix} \frac{\partial t_1}{\partial \tau} & \frac{\partial t_1}{\partial \theta} \\ \frac{\partial t_2}{\partial \tau} & \frac{\partial t_2}{\partial \theta} \end{bmatrix} \right) = \det \left(\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \right) = 1.\tag{5.45}$$

Astfel, ecuația (5.42) poate fi scrisă în noile coordonate astfel:

$$\overline{\left(\mu_T^{(k)}(\xi) - \bar{\xi} \right)^2} = \frac{1}{T^2} \iint_D K_\xi(\tau) d\tau d\theta,\tag{5.46}$$

unde D este domeniul bidimensional reprezentat în figura 5.2.(b). În ecuația (5.46), facem mai întâi integrarea după variabila θ . Se poate observa că limitele de variație ale lui θ depind de τ . Astfel, dacă $\tau \in [0, T]$ (corespunzător jumătății paralelogramului situate deasupra axei absciselor), $\theta \in [-\frac{T}{2}, \frac{T}{2} - \tau]$. Pentru $\tau \in [-T, 0]$, însă (jumătatea inferioară a paralelogramului D) avem $\theta \in [-\frac{T}{2} - \tau, \frac{T}{2}]$. În cele expuse anterior, s-a ținut cont că ecuațiile celor două drepte oblice ale paralelogramului D sunt $\theta + \tau = \frac{T}{2}$, respectiv

$\theta + \tau = -\frac{T}{2}$. Astfel, ecuația (5.46) devine:

$$\begin{aligned}
 \overline{\left(\mu_T^{(k)}(\xi) - \bar{\xi}\right)^2} &= \frac{1}{T^2} \left(\int_0^T \int_{-\frac{T}{2}}^{\frac{T}{2}-\tau} K_\xi(\tau) d\theta d\tau + \int_{-T}^0 \int_{-\frac{T}{2}-\tau}^{\frac{T}{2}} K_\xi(\tau) d\theta d\tau \right) \\
 &= \frac{1}{T^2} \left[\int_0^T K_\xi(\tau) \underbrace{\left(\int_{-\frac{T}{2}}^{\frac{T}{2}-\tau} d\theta \right)}_{T-\tau} d\tau + \int_{-T}^0 K_\xi(\tau) \underbrace{\left(\int_{-\frac{T}{2}-\tau}^{\frac{T}{2}} d\theta \right)}_{T+\tau} d\tau \right] \quad (5.47) \\
 &= \frac{1}{T^2} \left(\int_0^T (T-\tau) K_\xi(\tau) d\tau + \int_{-T}^0 (T+\tau) K_\xi(\tau) d\tau \right) \\
 &= \frac{1}{T^2} \int_{-T}^T (T-|\tau|) K_\xi(\tau) d\tau = \frac{1}{T} \int_{-T}^T \left(1 - \frac{|\tau|}{T}\right) K_\xi(\tau) d\tau.
 \end{aligned}$$

Ținând cont de relația (5.41), ajungem la forma finală a teoremei erodicității mediei,

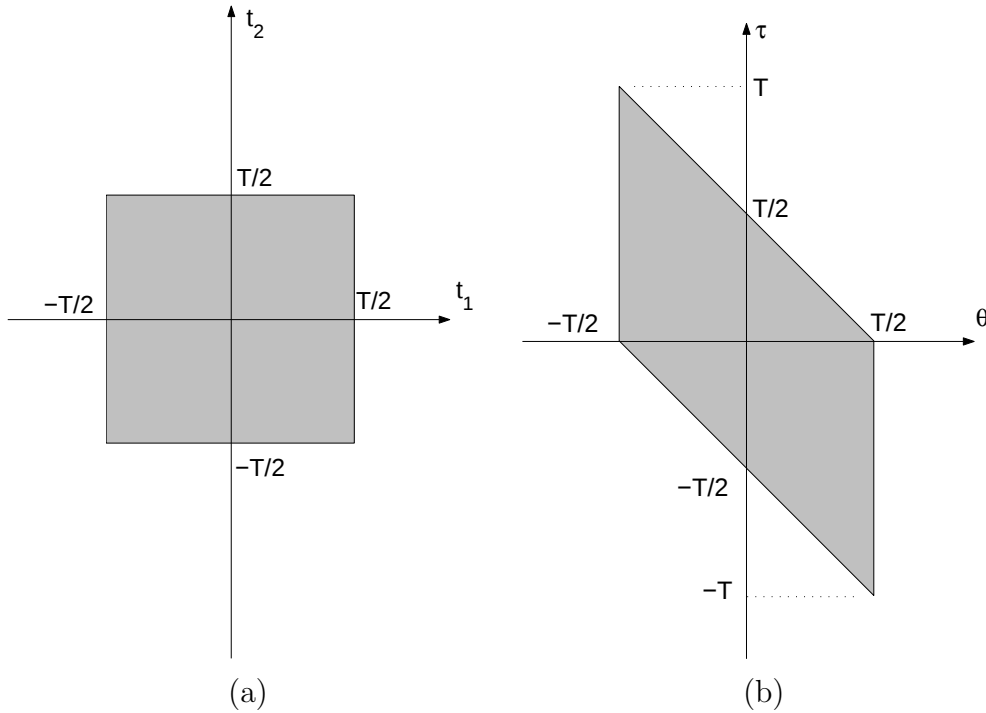


Figura 5.2: (a) Domeniul de integrare în coordonate (t_1, t_2) ; (b) Domeniul de integrare în coordonate (θ, τ)

care afirmă că un semnal este ergodic în sensul mediei dacă:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T}^T \left(1 - \frac{|\tau|}{T}\right) K_{\xi}(\tau) d\tau = 0. \quad (5.48)$$

Se observă că dacă semnalul este decorelat la infinit, adică dacă $\exists \tau_0 \in \mathbb{R}$ astfel încât $K_{\xi}(\tau) = 0$ pentru $\forall \tau \geq \tau_0$ (ipoteză cât se poate de verosimilă pentru majoritatea semnalelor reale, conform discuției purtate în contextul demonstrației proprietății 3 a funcției de autocorelație) atunci:

$$\left| \int_{-T}^T \left(1 - \frac{|\tau|}{T}\right) K_{\xi}(\tau) d\tau \right| = M < \infty, \quad \forall T \in \mathbb{R},$$

și, deci limita din (5.48) este 0, semnalul fiind ergodic în sensul mediei. Să mai precizăm că această ipoteză a decorelării a semnalului la infinit reprezintă o condiție suficientă, nu și necesară, relația (5.48) putând fi îndeplinită și de alte semnale (printre care și cele asimptotic decorelate).

Să mai remarcăm, în final, că demonstrația de mai sus se poate relua și pentru funcția de autocorelație, pornind, în loc de media (5.39) calculată nu pentru semnalul $\xi(t)$, ci pentru $\eta_{\tau}(t) = \xi(t)\xi(t + \tau)$:

$$\mu_T^{(k)}(\eta_{\tau}) = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \xi^{(k)}(t) \xi^{(k)}(t + \tau) dt. \quad (5.49)$$

Rezultatul final ar face, însă, să apară funcția de covariație a lui $\eta_{\tau}(t)$ care depinde de momente de ordin superior lui doi ale lui $\xi(t)$.

5.7 Densitatea spectrală de putere

În acest paragraf, ne propunem caracterizarea din punct de vedere spectral a semnalelor aleatoare staționare în sens larg.

Acest demers nu este unul trivial, și necesită o atenție sporită, datorită faptului că ipoteza de staționaritate a unui semnal implică faptul că acesta este de modul neintegrabil! Într-adevăr, dacă un semnal este staționar, atunci el este “obligat” să aibă suport spațial infinit ¹ (ceea ce implică faptul că, în realitate, nu există semnale pur staționare), și:

$$\int_{-\infty}^{\infty} |\xi(t)| dt = \infty. \quad (5.50)$$

¹Dacă nu ar fi așa, dacă semnalul ar avea un suport temporal finit, să zicem, $\xi(t) = 0$ pentru $t \notin [T_1, T_2]$, atunci mediile semnalului nu ar fi invariante în timp. De exemplu, $\overline{\xi^2(t_1)} = 0$ pentru $t_1 \notin [T_1, T_2]$, în timp ce $\overline{\xi^2(t_2)} \neq 0$ pentru $t_2 \in [T_1, T_2]$ și, deci, semnalul nu ar fi staționar.

Aceasta implică, mai departe, faptul ca nu putem calcula transformata Fourier a vreunei realizări particulare a semnalului, știut fiind că una din condițiile de existență a transformatei Fourier a unui semnal este integrabilitatea modulului acestuia.

Pentru a defini, totuși, o măsură care să caracterizeze spectral semnalul, vom recurge la următoarea construcție. Considerăm, mai întâi, o realizare particulară a semnalului, $\xi^{(k)}$, și considerăm, mai departe, o variantă trunchiată a acesteia, pe care o notăm $\xi_T^{(k)}$ și pe care o definim astfel:

$$\xi_T^{(k)}(t) = \begin{cases} \xi^{(k)}(t) & \text{dacă } |t| \leq \frac{T}{2} \\ 0 & \text{în rest} \end{cases}. \quad (5.51)$$

Datorită limitării suportului temporal, varianta trunchiată este de modul integrabil și, deci, transformata Fourier a acesteia există și o vom nota cu $X_T^{(k)}$:

$$X_T^{(k)}(\omega) = \mathcal{F} \left\{ \xi_T^{(k)}(t) \right\} (\omega) = \int_{-\infty}^{\infty} \xi_T^{(k)}(t) \exp(-j\omega t) dt. \quad (5.52)$$

Evident, semnalul original $\xi^{(k)}$ poate fi recuperat din echivalentul său spectral prin transformare Fourier inversă:

$$\xi_T^{(k)}(t) = \mathcal{F}^{-1} \left\{ X_T^{(k)}(\omega) \right\} (t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X_T^{(k)}(\omega) \exp(j\omega t) d\omega. \quad (5.53)$$

Energia realizării particulare trunchiate a semnalului este dată de:

$$\begin{aligned} E_{\xi_T^{(k)}} &= \int_{-\infty}^{\infty} \left(\xi_T^{(k)} \right)^2 dt = \int_{-\infty}^{\infty} \xi_T^{(k)} \xi_T^{(k)} dt \\ &= \int_{-\infty}^{\infty} \xi_T^{(k)} \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} X_T^{(k)}(\omega) \exp(j\omega t) d\omega \right) dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} X_T^{(k)}(\omega) \left(\int_{-\infty}^{\infty} \xi_T^{(k)} \exp(j\omega t) dt \right) d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} X_T^{(k)}(\omega) X_T^{(k)*}(\omega) d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| X_T^{(k)}(\omega) \right|^2 d\omega, \end{aligned} \quad (5.54)$$

unde prin $()^*$ am notat operatorul de conjugare complexă.

Puterea realizării particulare trunchiate a semnalului este dată de:

$$P_{\xi_T^{(k)}} = \frac{E_{\xi_T^{(k)}}}{T} = \frac{1}{2\pi T} \int_{-\infty}^{\infty} \left| X_T^{(k)}(\omega) \right|^2 d\omega. \quad (5.55)$$

Astfel, am reușit să definim o mărime, respectiv valoarea $\left|X_T^{(k)}(\omega)\right|^2$, care să reprezinte o caracteristică spectrală a semnalului. Ea, însă, caracterizează o singură realizare particulară, și aceea trunchiată, a acestuia. Pentru a obține o măsură care să caracterizeze semnalul aleator în ansamblul lui, și nu o realizare particulară oarecare a acestuia, trebuie să efectuăm o mediere statistică. Astfel, se definește puterea medie a semnalului trunchiat ξ_T ca fiind media statistică a puterii $P_{\xi_T^{(k)}}$:

$$P_{\xi_T} = \overline{P_{\xi_T^{(k)}}} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\overline{\left|X_T^{(k)}(\omega)\right|^2}}{T} d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\overline{\left|X_T^{(k)}(\omega)\right|^2}}{T} d\omega. \quad (5.56)$$

În sfârșit, ultima operațiune pe care o efectuăm este trecerea la limită când $T \rightarrow \infty$ a relației (5.56), pentru a elimina efectul trunchierii. Se definește, deci, puterea medie a semnalului netrunchiat ca fiind:

$$P_{\xi} = \lim_{T \rightarrow \infty} P_{\xi_T} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{T \rightarrow \infty} \frac{\overline{\left|X_T^{(k)}(\omega)\right|^2}}{T} d\omega. \quad (5.57)$$

Trebuie menționat că faptul că semnalul este de energie infinită este o consecință a duratei sale infinite; puterea lui este finită, deci limita de mai sus există!

Integrandul din ecuația (5.57) se notează cu q_{ξ} și se numește *densitatea spectrală de putere* a semnalului:

$$q_{\xi}(\omega) = \lim_{T \rightarrow \infty} \frac{\overline{\left|X_T^{(k)}(\omega)\right|^2}}{T}. \quad (5.58)$$

Se observă că puterea medie a semnalului se scrie ca

$$P_{\xi} = \frac{1}{2\pi} \int_{-\infty}^{\infty} q_{\xi}(\omega) d\omega, \quad (5.59)$$

ceea ce justifică întru totul denumirea dată mărimii q_{ξ} : din ecuația (5.59), se observă că semnificația cantității $q_{\xi}(\omega)d\omega$ este de *putere medie a semnalului conținută de armonicile acestuia din intervalul de frecvențe $[\omega, \omega + d\omega]$* .

Din relația (5.58), rezultă că densitatea spectrală de putere este o funcție *reală* și *pozitivă*, ceea ce este normal, având în vedere faptul că aceasta măsoară o putere, mărime prin definiție pozitivă. În plus, q_{ξ} este o funcție *pară*, această afirmație fiind justificată prin proprietatea transformatei Fourier a unui semnal real de a avea valori complex conjugate în puncte situate simetric față de 0.

În paragraful următor, vom enunța și demonstra o teoremă fundamentală a teoriei statistice a semnalelor, care stabilește legătura între mărimile statistice spectrale și temporale ce caracterizează semnalele aleatoare staționare.

5.8 Teorema Wiener–Hincin

Teorema Wiener–Hincin face legătura între funcția de autocorelație și densitatea spectrală de putere a semnalelor aleatoare staționare în sens larg.

Teoremă. Densitatea spectrală de putere a unui semnal aleator staționar în sens larg este transformata Fourier a funcției sale de autocorelație.

Demonstrație. Conform celor expuse mai înainte, densitatea spectrală de putere a semnalului poate fi scrisă ca:

$$q_\xi(\omega) = \lim_{T \rightarrow \infty} \frac{\overline{X_T^{(k)}(\omega)^2}}{T} = \lim_{T \rightarrow \infty} \frac{\overline{X_T^{(k)}(\omega) X_T^{(k)*}(\omega)}}{T}, \quad (5.60)$$

Pornind de la (5.52), avem:

$$X_T^{(k)}(\omega) = \int_{-\infty}^{\infty} \xi_T^{(k)}(t_1) \exp(-j\omega t_1) dt_1 = \int_{-\frac{T}{2}}^{\frac{T}{2}} \xi^{(k)}(t_1) \exp(-j\omega t_1) dt_1, \quad (5.61)$$

și observând că

$$X_T^{(k)*}(\omega) = \int_{-\frac{T}{2}}^{\frac{T}{2}} \xi^{(k)}(t_2) \exp(j\omega t_2) dt_2, \quad (5.62)$$

rezultă că ecuația (5.60) poate fi scrisă ca:

$$\begin{aligned} q_\xi(\omega) &= \lim_{T \rightarrow \infty} \frac{1}{T} \overline{\int_{-\frac{T}{2}}^{\frac{T}{2}} \int_{-\frac{T}{2}}^{\frac{T}{2}} \xi_T^{(k)}(t_1) \xi_T^{(k)}(t_2) \exp(-j\omega(t_1 - t_2)) dt_1 dt_2} \\ &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \int_{-\frac{T}{2}}^{\frac{T}{2}} \underbrace{\xi_T^{(k)}(t_1) \xi_T^{(k)}(t_2)}_{R_\xi(t_1, t_2)} \exp(-j\omega(t_1 - t_2)) dt_1 dt_2 \\ &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \int_{-\frac{T}{2}}^{\frac{T}{2}} R_\xi(t_1 - t_2) \exp(-j\omega(t_1 - t_2)) dt_1 dt_2 \end{aligned} \quad (5.63)$$

Reluând pașii demonstrației teoremei ergodicității mediei din paragraful 5.6.3, mai precis trecerea de la integrala bidimensională din (5.42) la cea unidimensională din (5.47), rezultă că, pornind de la (5.63), putem scrie densitatea spectrală de putere ca:

$$\begin{aligned}
q_\xi(\omega) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T}^T (T - |\tau|) R_\xi(\tau) \exp(-j\omega\tau) d\tau \\
&= \lim_{T \rightarrow \infty} \int_{-T}^T R_\xi(\tau) \exp(-j\omega\tau) d\tau - \lim_{T \rightarrow \infty} \int_{-T}^T \frac{|\tau|}{T} R_\xi(\tau) \exp(-j\omega\tau) d\tau \\
&= \int_{-\infty}^{\infty} R_\xi(\tau) \exp(-j\omega\tau) d\tau - \lim_{T \rightarrow \infty} \int_{-T}^T \frac{|\tau|}{T} R_\xi(\tau) \exp(-j\omega\tau) d\tau
\end{aligned} \tag{5.64}$$

Se poate observa cu ușurință faptul că primul termen din ecuația (5.64) reprezintă exact transformata Fourier a funcției de autocorelație a semnalului $\xi(t)$. Pentru ca demonstrația să fie completă, mai trebuie, deci, demonstrat că al doilea termen al ecuației (5.64) este nul, cu alte cuvinte, că:

$$\lim_{T \rightarrow \infty} \int_{-T}^T \frac{|\tau|}{T} R_\xi(\tau) \exp(-j\omega\tau) d\tau = 0. \tag{5.65}$$

Pentru aceasta, trebuie să folosim ipoteza de integrabilitate a modulului funcției de autocorelație a semnalului, ipoteză în absența căreia nici nu se pune problema calculului transformatei Fourier a acesteia (v. discuția purtată în paragraful 5.7). Vom presupune, deci, că:

$$\int_{-\infty}^{\infty} |R_\xi(\tau)| d\tau = M < \infty. \tag{5.66}$$

Se poate arăta că din condiția de integrabilitate a modulului funcției de autocorelație din ecuația (5.66) rezultă că:

$$\lim_{T \rightarrow \infty} \int_T^\infty |R_\xi(\tau)| d\tau = 0. \tag{5.67}$$

Într-adevăr, valoarea integralei din ecuația (5.67) este aria A_T din figura 5.3, care obligatoriu trebuie să tindă la zero când $T \rightarrow \infty$ pentru ca întreaga arie de sub graficul funcției să fie finită. Ecuația (5.67) poate fi rescrisă:

$$\forall \varepsilon > 0, \exists T_\varepsilon, \text{ astfel încât } \int_{T_\varepsilon}^\infty |R_\xi(\tau)| d\tau < \varepsilon. \tag{5.68}$$

Fie, așadar, $\varepsilon > 0$, și fie T_ε (fixat) astfel încât să fie satisfăcută relația (5.68). Mai întâi, facem următoarea observație:

$$\left| \int_{-T}^T \frac{|\tau|}{T} R_\xi(\tau) \exp(-j\omega\tau) d\tau \right| \leq \int_{-T}^T \left| \frac{|\tau|}{T} R_\xi(\tau) \exp(-j\omega\tau) \right| d\tau = \int_{-T}^T \frac{|\tau|}{T} |R_\xi(\tau)| d\tau \tag{5.69}$$

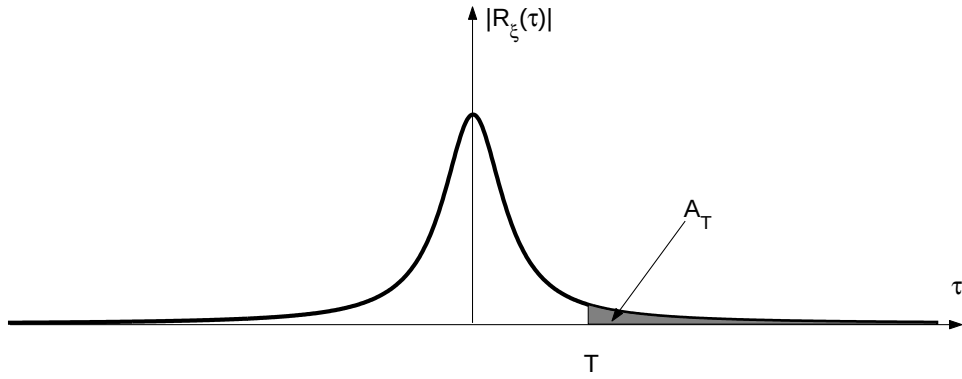


Figura 5.3: O funcție de modul integrabil.

Apoi, observăm că pentru T suficient de mare ($T > T_\varepsilon$), putem scrie:

$$\int_{-T}^T \frac{|\tau|}{T} |R_\xi(\tau)| d\tau = \int_{-T}^{-T_\varepsilon} \dots d\tau + \int_{-T_\varepsilon}^{T_\varepsilon} \dots d\tau + \int_{T_\varepsilon}^T \dots d\tau. \quad (5.70)$$

În relația de mai sus, integranzii celor trei integrale din partea dreapta sunt identici cu cel din stânga, și au fost omiși din motive de simplitate a scrierii. În continuare, vom considera separat fiecare din cele trei integrale din relația de mai sus, și vom arăta că fiecare dintre ele poate fi făcută mai mică decât ε . Astfel, pentru termenul din mijloc putem scrie:

$$\int_{-T_\varepsilon}^{T_\varepsilon} \frac{|\tau|}{T} |R_\xi(\tau)| d\tau = \frac{1}{T} \underbrace{\int_{-T_\varepsilon}^{T_\varepsilon} |\tau| |R_\xi(\tau)| d\tau}_{\text{constant}} < \varepsilon \text{ pt. } T \text{ suf. de mare} \quad (5.71)$$

În relația (5.71) s-a folosit faptul că integrala este una pe un domeniu finit dintr-o funcție finită, și, deci, este constantă. În continuare, pentru al treilea termen, avem:

$$\int_{T_\varepsilon}^T \frac{|\tau|}{T} |R_\xi(\tau)| d\tau < \int_{T_\varepsilon}^T |R_\xi(\tau)| d\tau < \int_{T_\varepsilon}^{\infty} |R_\xi(\tau)| d\tau < \varepsilon, \quad (5.72)$$

și, în mod absolut similar cu relația (5.72), se poate arăta că:

$$\int_{-T}^{-T_\varepsilon} \frac{|\tau|}{T} |R_\xi(\tau)| d\tau < \dots < \varepsilon. \quad (5.73)$$

În concluzie, introducând relațiile (5.71) (5.72) și (5.73) în ecuația (5.70), și, suplimentar, ținând cont de relația (5.69), putem afirma că pentru $\forall \varepsilon > 0$, $\exists T$ suficient de mare astfel

încât:

$$\left| \int_{-T}^T \frac{|\tau|}{T} R_{\xi}(\tau) \exp(-j\omega\tau) d\tau \right| < 3\varepsilon, \quad (5.74)$$

ceea ce este echivalent cu a spune că relația (5.65) este demonstrată, ceea ce, mai departe, face ca relația (5.64) să devină:

$$q_{\xi}(\omega) = \int_{-\infty}^{\infty} R_{\xi}(\tau) \exp(-j\omega\tau) d\tau. \quad (5.75)$$

Cu alte cuvinte, demonstrația teoremei Wiener–Hincin, care afirmă că densitatea spectrală de putere și funcția de autocorelație a semnalelor aleatoare staționare în sens larg fac pereche Fourier, este terminată!

5.9 Densitatea spectrală de putere de interacțiune

În mod similar cu modul de definire al densității spectrale de putere a unui semnal, se poate defini densitatea spectrală de putere de interacțiune între două semnale aleatoare staționare $\xi(t)$ și $\eta(t)$, după relația:

$$q_{\xi\eta}(\omega) = \lim_{T \rightarrow \infty} \frac{\overline{X_T^{(k)*}(\omega) Y_T^{(k)}(\omega)}}{T} \quad (5.76)$$

Din relația (5.76) este evident că se poate defini similar și densitatea reciprocă $q_{\eta\xi}$ care este legată de $q_{\xi\eta}$ prin:

$$q_{\eta\xi}(\omega) = q_{\xi\eta}^*(\omega) \quad (5.77)$$

Densitatea de putere de interacțiune între două semnale nu mai are semnificația fizică de putere propriu-zisă, și este denumită astfel numai prin analogie cu densitatea de putere a unui semnal (dealtfel, se poate observa cu ușurință că $q_{\xi\eta}(\omega)$ este o funcție cu valori complexe). Densitatea de interacțiune este mai mult o măsură a dependenței statistice dintre cele două semnale, exprimată în frecvență. Se poate arăta, reluând demonstrația din paragraful 5.8, că densitatea spectrală de putere de interacțiune este transformata Fourier a funcției de intercorelație dintre cele două semnale:

$$q_{\xi\eta}(\omega) = \mathcal{F}\{R_{\xi\eta}(\tau)\}(\omega) = \int_{-\infty}^{\infty} R_{\xi\eta}(\tau) \exp(-j\omega\tau) d\tau \quad (5.78)$$

5.10 Zgomotul alb

Un semnal aleator staționar se numește zgomot de bandă largă dacă densitatea sa spectrală de putere poate fi considerată cu bună aproximație constantă într-o bandă largă de

frecvențe. Se poate, deci, scrie:

$$q_{\xi}(\omega) \approx \begin{cases} K & \text{dacă } |\omega| \leq \omega_T \\ 0 & \text{în rest} \end{cases}. \quad (5.79)$$

unde ω_T este o frecvență foarte mare (mult mai mare decât frecvențele celorlalte semnale considerate în aplicația respectivă). În teorie, acest tip de semnal este *modelat* cu un semnal ce are densitatea spectrală de putere constantă în toată gama de frecvențe. Acest semnal se numește *zgomot alb*, denumire ce provine din analogia cu lumina albă, care are componente în toată gama de frecvențe vizibile.

Deci, zgomotul alb are

$$q_{\xi}(\omega) = K, \quad \forall \omega, \quad (5.80)$$

și, drept urmare, are funcția de autocorelație

$$R_{\xi}(\tau) = \mathcal{F}^{-1}\{q_{\xi}(\omega)\}(\tau) = K\delta(\tau). \quad (5.81)$$

Funcția de autocorelație și densitatea spectrală de putere ale zgomotului alb sunt ilustrate în figura 5.4.

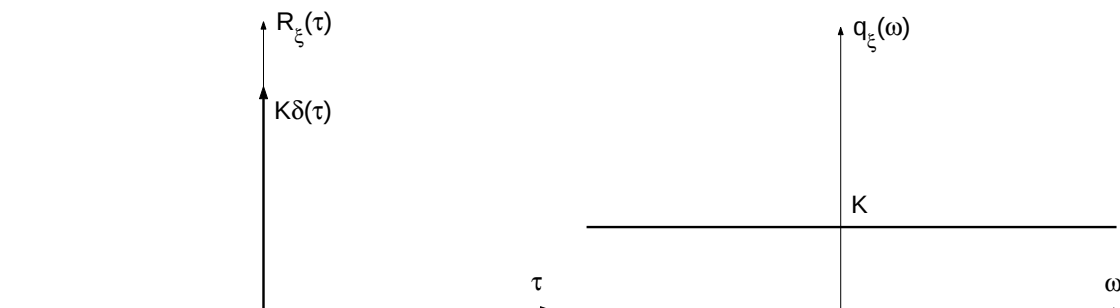


Figura 5.4: Funcția de autocorelație și densitatea spectrală de putere ale zgomotului alb.

Din (5.81), rezultă că zgomotul alb este un semnal aleator *pur decorelat*: întrucât $R_{\xi}(t_1) = 0 \quad \forall t_1 \neq 0$, rezultă că oricare două valori ale semnalului, oricât de apropiate în timp una de cealaltă, sunt decorelate. În același timp, din (5.80), rezultă că el este de putere infinită. Evident, acest semnal este o idealizare, nu poate exista în realitate, dar servește cu bună aproximație de model pentru multe din semnalele perturbatoare reale.

5.11 Trecerea semnalelor aleatoare prin sisteme liniare invariante în timp

În acest paragraf vom aborda problema filtrării liniare a semnalelor aleatoare staționare. Sistemele liniare invariante în timp (pe scurt, filtrele liniare) sunt de o importanță aparte

în studiul prelucrării de semnale, fiind de departe subiectul cel mai studiat în acest domeniu. Este de interes, deci, studierea comportamentului semnalelor aleatoare la trecerea prin astfel de filtre. Mai precis, ne interesează să găsim legătura între mărimile statistice ale semnalelor de la intrarea și ieșirea filtrelor. Pentru început, în paragraful următor, vom face o scurtă trecere în revistă a noțiunilor de bază referitoare la sistemele liniare, invariante în timp.

5.11.1 Sisteme liniare invariante în timp

Un sistem se numește *liniar* dacă, aplicând la intrare o combinație liniară a două semnale $ax_1(t) + bx_2(t)$ cu $a, b \in \mathbb{R}$, semnalul de ieșire poate fi scris sub forma $ay_1(t) + by_2(t)$, unde $y_1(t)$ și $y_2(t)$ sunt răspunsurile sistemului la $x_1(t)$, respectiv $x_2(t)$. (v. figura 5.5).

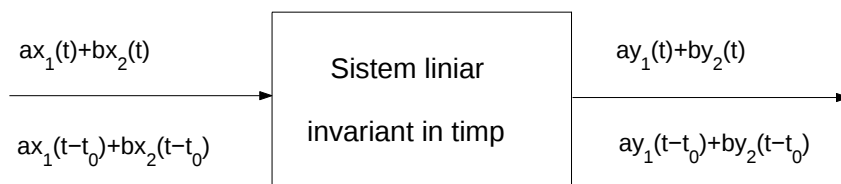


Figura 5.5: Sistem liniar invariant în timp.

Sistemul se numește *invariant în timp* dacă acțiunea sa asupra unui semnal $x(t)$ este aceeași indiferent de momentul când este aplicat semnalul respectiv la intrarea sistemului.

Se știe din teoria sistemelor că sistemele liniare invariante în timp (pe care le vom numi, pe scurt, *filtre liniare*) au ca funcții proprii semnalele sinusoidale. Altfel spus, dacă aplicăm la intrarea unui filtru liniar un semnal sinusoidal de frecvență ω_0 , semnalul de ieșire va fi tot o sinusoidă de frecvență ω_0 (cu amplitudinea și faza modificate față de sinusoida de intrare). Un filtru liniar va fi, deci, complet caracterizat de o funcție care să descrie modificarea amplitudinii și a fazei sinusoidelor pentru fiecare frecvență ω_0 , funcție care se numește *funcție de transfer* și se notează $H(\omega)$.

Spre exemplu, să considerăm filtrul liniar format dintr-o celulă RC prezentat în figura 5.6. Dacă tensiunea de intrare este sinusoidală, $x(t) = X \cos(\omega_0 t + \varphi_X)$, atunci putem scrie tensiunea de ieșire sub forma $y(t) = Y \cos(\omega_0 t + \varphi_Y)$. Relația dintre amplitudinile și fazele semnalelor de intrare și ieșire poate fi exprimată simplu, considerând reprezentarea semnalelor sub formă complexă. Astfel, dacă avem $\underline{X} = X \exp(j\varphi_X)$ și $\underline{Y} = Y \exp(j\varphi_Y)$, atunci putem scrie:

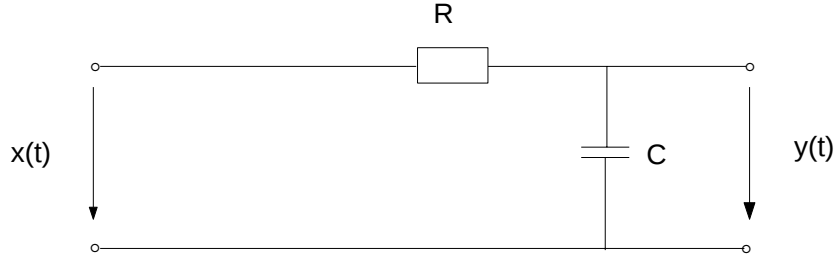
$$\underline{Y} = \frac{\frac{1}{j\omega_0 C}}{R + \frac{1}{j\omega_0 C}} \underline{X} = \frac{1}{1 + j\omega_0 RC} \underline{X}.$$

Cantitatea $\frac{1}{1 + j\omega_0 RC}$ este exact valoarea funcției de transfer a filtrului pentru frecvența ω_0 :

$$H(\omega_0) \stackrel{\text{not}}{=} \frac{1}{1 + j\omega_0 RC}.$$

Putem, deci, scrie că:

$$\underline{Y} = H(\omega_0) \underline{X}. \quad (5.82)$$

**Figura 5.6:** Celulă RC.

În cazul în care semnalul $x(t)$ aplicat la intrarea filtrului liniar nu este sinusoidal, putem explica forma semnalului de ieșire $y(t)$ tot prin reducere la cazul sinusoidal. Acest lucru este posibil întrucât știm, de la analiza Fourier a semnalelor, că orice semnal de modul integrabil $x(t)$ poate fi descompus ca o sumă (infinită) de semnale pur sinusoidale. Într-adevăr, considerând transformata Fourier a lui $x(t)$:

$$X(\omega) = \mathcal{F}\{x(t)\}(\omega) = \int_{-\infty}^{\infty} x(t) \exp(-j\omega t) dt, \quad (5.83)$$

atunci îl putem scrie pe $x(t)$ sub forma:

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) \exp(j\omega t) d\omega. \quad (5.84)$$

Relația (5.84) reprezintă exact descompunerea dorită a semnalului $x(t)$ într-o sumă de sinusoidale pure.

Ținând cont de faptul că, în ipoteza în care $x(t)$ este un semnal real, avem $X(-\omega) = X^*(\omega) = |X(\omega)| \exp(-j\varphi(\omega))$, atunci relația (5.84) poate fi adusă la forma:

$$x(t) = \frac{1}{\pi} \int_0^{\infty} |X(\omega)| \cos(\omega t + \varphi(\omega)) d\omega, \quad (5.85)$$

ceea ce justifică afirmația de mai sus. Modulul $|X(\omega)|$ și faza $\varphi(\omega)$ ale cantității complexe $X(\omega)$ sunt amplitudinea, respectiv faza cosinusoidei de frecvență ω ce intră în componența semnalului $x(t)$. Cu alte cuvinte, cantitatea $X(\omega)$ este tocmai reprezentarea sub formă complexă a sinusoidelor respective.

Apoi, pentru *fiecare* componentă sinusoidală de frecvență ω din semnal, putem calcula faza și amplitudinea la ieșire în funcție de cele la intrare (vezi relația (5.82)) precum:

$$Y(\omega) = H(\omega)X(\omega) \quad \forall \omega, \quad (5.86)$$

după care “recompunem” semnalul de ieșire sumând toate sinusoidalele $Y(\omega)$:

$$y(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} Y(\omega) \exp(j\omega t) d\omega. \quad (5.87)$$

Ținând cont de proprietatea transformatei Fourier privind produsul de convoluție, relația (5.86) poate fi scrisă direct în domeniul temporal, considerând transformata Fourier inversă a funcției de transfer, notată $h(t)$ și numită *funcție pondere*:

$$h(t) = \mathcal{F}^{-1}\{H(\omega)\}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} H(\omega) \exp(j\omega t) d\omega. \quad (5.88)$$

Astfel, relația (5.86) poate fi scrisă ca:

$$y(t) = h(t) \star x(t) = \int_{-\infty}^{\infty} h(\tau) x(t - \tau) d\tau = \int_{-\infty}^{\infty} h(t - \tau) x(\tau) d\tau. \quad (5.89)$$

În paragraful următor vom arăta cum influențează trecerea prin filtre liniare mărimile statistice ce caracterizează semnalele aleatoare.

5.11.2 Relații între mărimile statistice la trecerea prin filtre liniare

Considerând că la intrarea unui filtru liniar aplicăm un semnal aleator staționar $\xi(t)$, la ieșire vom obține un semnal aleator, pe care-l vom nota cu $\eta(t)$. Se poate arăta cu ușurință că și semnalul de la ieșire este staționar.

În continuare, vom determina legătura între mărimile statistice ale celor două semnale. Rescriind relația (5.89), obținem pentru orice realizare particulară a acestora:

$$\eta(t) = h(t) \star \xi(t) = \int_{-\infty}^{\infty} h(\tau) \xi(t - \tau) d\tau. \quad (5.90)$$

Înmulțind relația (5.90) cu $\xi(t - \theta)$ și aplicând operația de mediere statistică, obținem:

$$\overline{\eta(t)\xi(t - \theta)} = \overline{\int_{-\infty}^{\infty} h(\tau) \xi(t - \tau) \xi(t - \theta) d\tau} = \int_{-\infty}^{\infty} h(\tau) \overline{\xi(t - \tau)\xi(t - \theta)} d\tau. \quad (5.91)$$

Prin identificare de termeni, ținând cont de ecuațiile (5.31) și (5.21), și de faptul că funcția de autocorelație a unui semnal este pară, rezultă:

$$R_{\xi\eta}(\theta) = \int_{-\infty}^{\infty} h(\tau) R_{\xi}(\theta - \tau) d\tau = h(\theta) \star R_{\xi}(\theta), \quad (5.92)$$

ceea ce, în domeniul frecvențial, devine:

$$q_{\xi\eta}(\omega) = H(\omega)q_{\xi}(\omega). \quad (5.93)$$

Pe de altă parte, dacă înmulțim relația (5.90) cu $\eta(t - \theta)$, obținem:

$$\overline{\eta(t)\eta(t - \theta)} = \overline{\int_{-\infty}^{\infty} h(\tau)\xi(t - \tau)\eta(t - \theta)d\tau} = \int_{-\infty}^{\infty} h(\tau)\overline{\xi(t - \tau)\eta(t - \theta)}d\tau, \quad (5.94)$$

ceea ce, prin identificare de termeni, ne dă:

$$R_{\eta}(\theta) = \int_{-\infty}^{\infty} h(\tau)R_{\xi\eta}(\tau - \theta)d\tau = \int_{-\infty}^{\infty} h(\tau)R_{\eta\xi}(\theta - \tau)d\tau = h(\theta) \star R_{\eta\xi}(\theta). \quad (5.95)$$

Rescriind relația (5.95) în frecvență, obținem:

$$q_{\eta}(\omega) = H(\omega)q_{\eta\xi}(\omega). \quad (5.96)$$

Folosind în continuare relația (5.77), și înlocuind pe $q_{\xi\eta}$ cu valoarea lui dată de relația (5.93), avem:

$$q_{\eta}(\omega) = H(\omega)q_{\xi\eta}^*(\omega) = H(\omega)H^*(\omega)q_{\xi}^*(\omega) = |H(\omega)|^2q_{\xi}^*(\omega), \quad (5.97)$$

de unde, ținând cont de faptul că densitatea spectrală de putere a unui semnal e o funcție reală, rezultă:

$$q_{\eta}(\omega) = |H(\omega)|^2q_{\xi}(\omega). \quad (5.98)$$

Relația (5.98) este cea căutată, care stabilește legătura între mărimile statistice ale semnalelor aleatoare aflate la intrarea și ieșirea filtrului liniar. Evident, funcția de autocorelație a semnalului de la ieșire poate fi scrisă:

$$R_{\eta}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} q_{\eta}(\omega) \exp(j\omega\tau)d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} |H(\omega)|^2q_{\xi}(\omega) \exp(j\omega\tau)d\omega \quad (5.99)$$

În continuare, vom prezenta câteva exemple de filtre liniare.

5.11.3 Trecerea semnalelor aleatoare prin filtrul trece-jos ideal

Filtrul trece-jos (FTJ) ideal este acel filtru care lasă să treacă nealterate toate armonicile semnalului cu frecvențe mai mici decât un anumit prag, în timp ce armonicile cu frecvențe superioare pragului respectiv sunt complet rejectate. Un asemenea comportament este modelat cu ajutorul unei funcții de transfer $H(\omega) = |H(\omega)| \exp(j\phi(\omega))$ ale cărei modul și fază sunt date de (v. figura 5.7):

$$|H(\omega)| = \begin{cases} A & \text{dacă } |\omega| \leq \omega_0 \\ 0 & \text{în rest} \end{cases}, \quad (5.100)$$

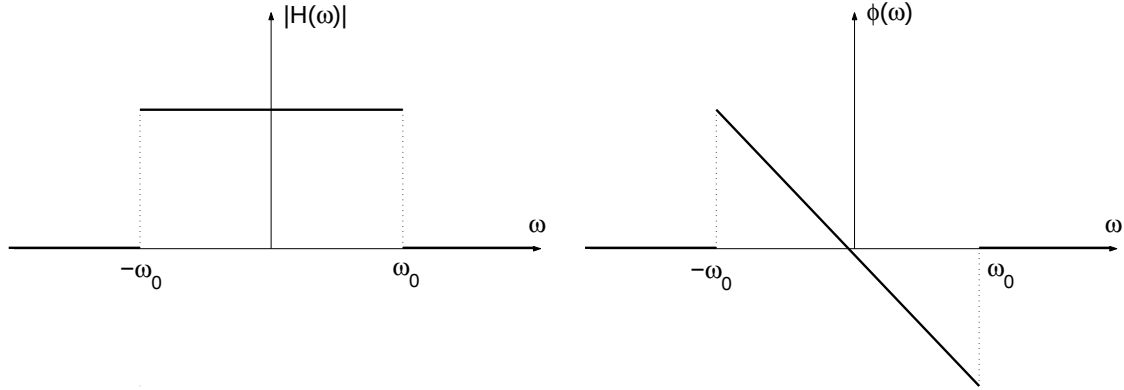


Figura 5.7: Modulul și faza unui filtru trece-jos ideal.

respectiv,

$$\phi(\omega) = \begin{cases} -\omega t_0 & \text{dacă } |\omega| \leq \omega_0 \\ 0 & \text{în rest} \end{cases}, \quad (5.101)$$

Frecvența ω_0 se numește frecvență de tăiere a filtrului și reprezintă frecvența maximă din semnalul de intrare care este lăsată să treacă de filtru, iar t_0 se numește timp de întârziere de grup, și reprezintă întârzierea indusă de filtru în semnal.

Funcția pondere a filtrului $h(t)$ se obține aplicând transformata Fourier inversă funcției de transfer:

$$\begin{aligned} h(t) &= \mathcal{F}^{-1}\{H(\omega)\}(t) = \frac{1}{2\pi} \int_{-\omega_0}^{\omega_0} A \exp(-j\omega t_0) \exp(j\omega t) d\omega \\ &= \frac{A}{2\pi} \frac{1}{j(t-t_0)} \exp(j\omega(t-t_0)) \Big|_{-\omega_0}^{\omega_0} = \frac{A}{2\pi} \frac{1}{j(t-t_0)} 2j \sin(\omega_0(t-t_0)) \\ &= \frac{A\omega_0}{\pi} \text{sinc}(\omega_0(t-t_0)). \end{aligned} \quad (5.102)$$

unde am folosit relația $\sin(x) = \frac{\exp(jx) - \exp(-jx)}{2j}$, iar $\text{sinc}(x)$ este funcția sinus cardinal, definită ca:

$$\text{sinc}(x) \triangleq \frac{\sin(x)}{x}. \quad (5.103)$$

Se observă că $h(t) \neq 0$ pentru $t < 0$, de unde rezultă că filtrul nu este cauzal², și, deci, nu poate fi implementat în timp real.

Dacă la intrarea filtrului se aplică un semnal aleator staționar $\xi(t)$ având densitatea spectrală de putere $q_\xi(\omega)$, atunci densitatea spectrală de putere a semnalului la ieșire $\eta(t)$ se scrie, ținând cont de (5.98), ca:

$$q_\eta(\omega) = \begin{cases} A^2 q_\xi(\omega) & \text{dacă } |\omega| \leq \omega_0 \\ 0 & \text{în rest} \end{cases}, \quad (5.104)$$

²Un filtru se numește cauzal dacă semnalul de ieșire apare ca rezultat al aplicării semnalului de intrare la intrarea filtrului. Aceasta este echivalentă cu condiția ca $h(t) = 0 \forall t < 0$. Într-adevăr, ținând cont de prima egalitate din relația (5.89), observăm că dacă $\exists \tau_1 < 0$ pentru care $h(\tau_1) \neq 0$, atunci în expresia lui $y(t)$ va apărea $x(t - \tau_1)$ care este valoarea lui x la un moment de timp ulterior lui t .

în timp ce funcția lui de autocorelație devine:

$$R_\eta(\tau) = \frac{A^2}{2\pi} \int_{-\omega_0}^{\omega_0} q_\xi(\omega) \exp(j\omega\tau) d\omega. \quad (5.105)$$

În cazul particular în care semnalul de intrare este zgomot alb, cu densitatea spectrală de putere dată de relația (5.80), atunci funcția de autocorelație a semnalului la ieșire devine:

$$R_\eta(\tau) = \frac{A^2}{2\pi} \int_{-\omega_0}^{\omega_0} K \exp(j\omega\tau) d\omega = \frac{A^2 K \omega_0}{\pi} \text{sinc}(\omega_0 \tau). \quad (5.106)$$

În obținerea rezultatului de mai sus am reluat calculul de la (5.102). Forma funcției de autocorelație a semnalului de la ieșirea filtrului trece-jos ideal este prezentată în figura 5.8.

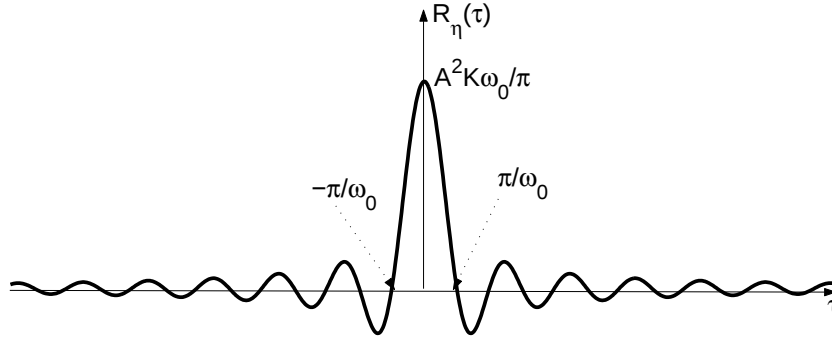


Figura 5.8: Funcția de autocorelație a semnalului de la ieșirea filtrului trece-jos ideal având la intrare un zgomot alb.

Este de remarcat aici un aspect important, și anume acela că, deși semnalul la intrare este complet decorelat, prin filtrarea acestuia cu un FTJ, în el se induce o corelație. Mai mult, considerând într-o primă aproximație, că numai lobul principal al funcției “sinc”, situat în intervalul $[-\frac{\pi}{\omega_0}, \frac{\pi}{\omega_0}]$, are valori importante, și neglijând restul funcției, putem afirma că această corelație este cu atât mai importantă cu cât banda de trecere a filtrului (în cazul de față ω_0) este mai îngustă. Această concluzie, după cum vom vedea în paragraful următor, este valabilă pentru orice fel de caracteristică frecvențială a filtrului.

Inducerea corelației în semnal este ilustrată în figurile 5.9, 5.10, și 5.11. În figura 5.9 sunt prezentate versiunile originală $\xi(t)$ și filtrate cu filtre trece-jos cu frecvență de tăiere medie $\eta(t)$, respectiv joasă $\zeta(t)$ a unui semnal zgomot alb. Se poate observa că pe măsură ce frecvența de tăiere scade, cu alte cuvinte, selectivitatea filtrului crește, semnalul este mai neted, ceea ce înseamnă că este mai corelat. Acest fapt reiese și mai evident din figura 5.10, care prezintă norul de puncte bidimensional $(s(t); s(t + \tau))$ pentru $\tau = 1$ pentru fiecare dintre cele trei semnale³. Se observă faptul că semnalul original este perfect

³S-a presupus că semnalele sunt ergodice în sensul autocorelației.

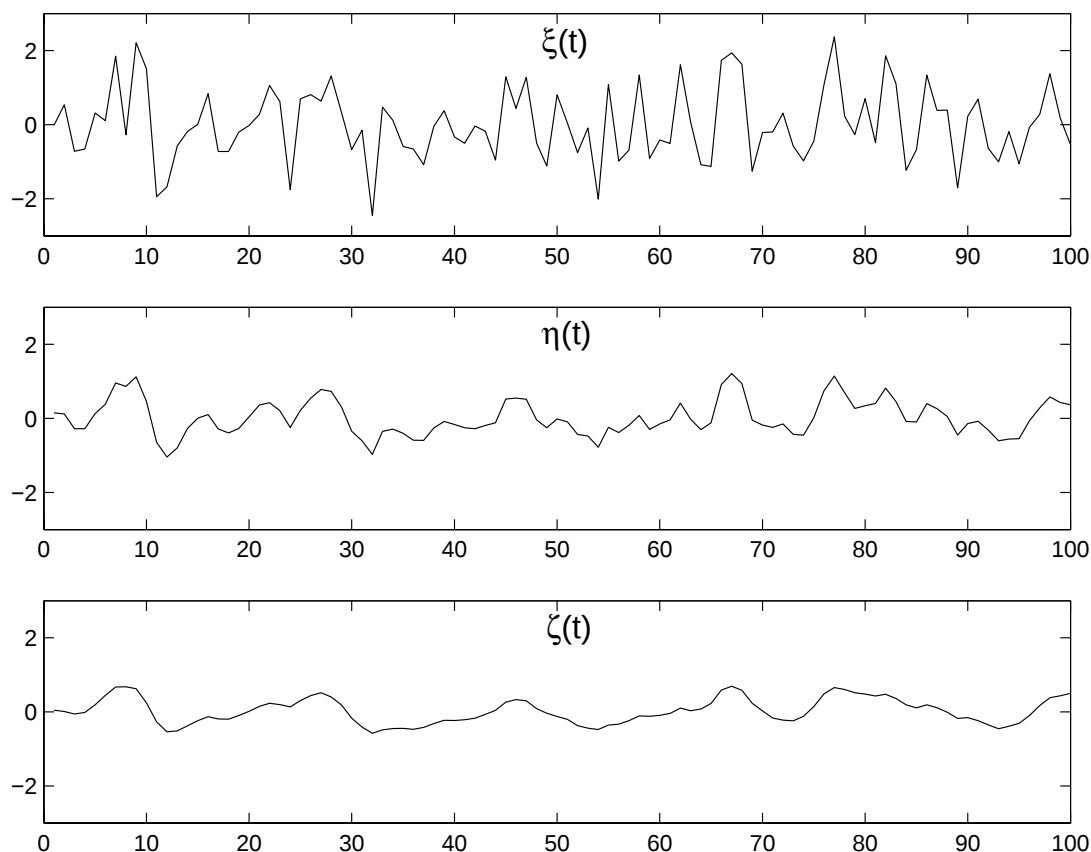


Figura 5.9: Realizare particulară a unui semnal zgomot alb, notat $\xi(t)$ și versiunile sale filtrate cu filtre trece-jos cu frecvență de tăiere ω_0 medie, notat $\eta(t)$, respectiv joasă, notat $\zeta(t)$.

decorelat (norul de puncte este perfect haotic) în timp ce între eșantioanele semnalelor filtrate există o corelație, fapt indicat de forma liniară a norului de puncte. Mai mult, corelația din semnalul filtrat cu filtrul mai selectiv $\zeta(t)$ (figura 5.10.(c)) este mai mare decât cea indusă de filtrul mai puțin selectiv $\eta(t)$ (figura 5.10.(b)).

În figura 5.11 este prezentat norul de puncte bidimensional $(s(t), s(t + \tau))$ pentru cele două semnale filtrate ($\eta(t)$ și $\zeta(t)$) pentru diverse valori ale lui τ . Corelația indusă în $\zeta(t)$ este mai puternică: valori relativ distanțate în timp ale lui $\zeta(t)$ încă mai sunt corelate, în timp ce corelația se “stinge” mai repede în $\eta(t)$.

5.11.4 Trecerea semnalelor aleatoare prin filtrul trece-bandă ideal

Filtrul trece-bandă (FTB) ideal este filtrul care lasă să treacă nealterate armonicile semnalului de intrare situate într-un interval de frecvențe dat și elimină complet celelalte

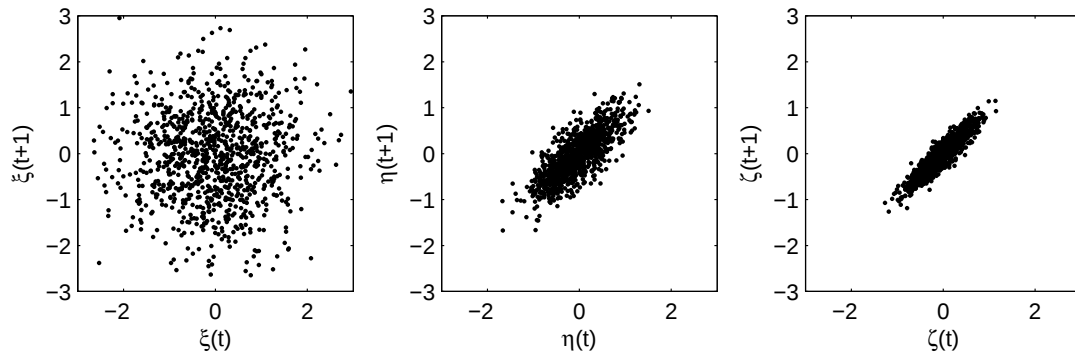


Figura 5.10: Norul de puncte $(s(t), s(t+1))$ pentru semnalele din figura 5.9.

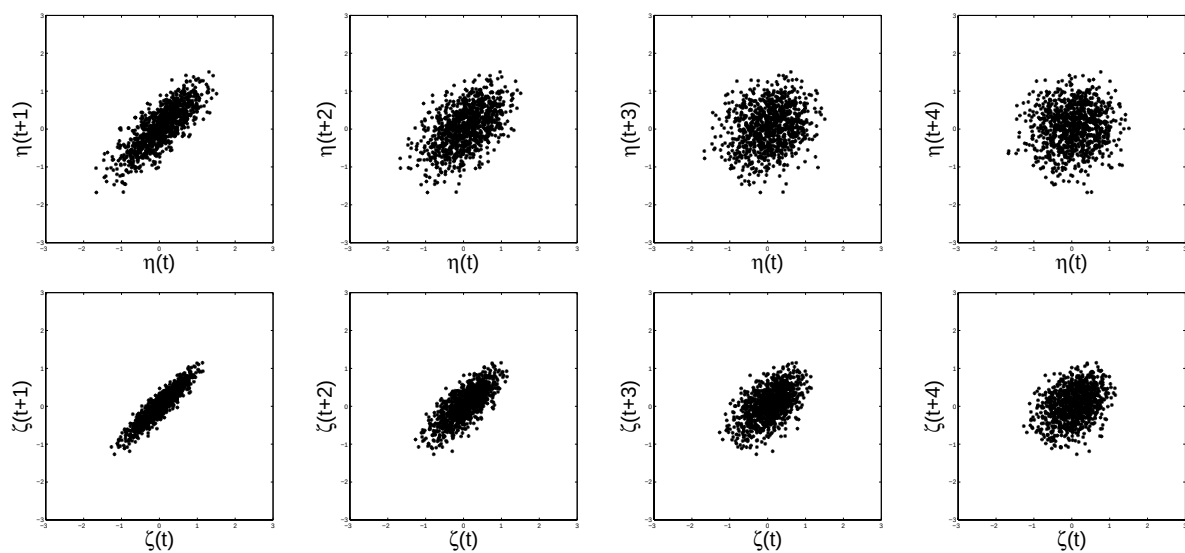


Figura 5.11: Norul de puncte $(s(t), s(t+\tau))$ pentru semnalul filtrat cu FTJ mai puțin selectiv (rândul de sus), respectiv pentru filtrul mai selectiv (rândul de jos) pentru $\tau = 1, 2, 3, 4$.

armonici. Modulul și faza funcției de transfer $H(\omega)$ sunt date de:

$$|H(\omega)| = \begin{cases} A & \text{dacă } |\omega| \in [\omega_0 - \frac{\Delta\omega}{2}, \omega_0 + \frac{\Delta\omega}{2}] \\ 0 & \text{în rest} \end{cases} \quad (5.107)$$

$$|\phi(\omega)| = \begin{cases} -(\omega - \omega_0)t_0 & \text{dacă } \omega \in [\omega_0 - \frac{\Delta\omega}{2}, \omega_0 + \frac{\Delta\omega}{2}] \\ -(\omega + \omega_0)t_0 & \text{dacă } \omega \in [-\omega_0 - \frac{\Delta\omega}{2}, -\omega_0 + \frac{\Delta\omega}{2}] \\ 0 & \text{în rest} \end{cases} \quad (5.108)$$

unde ω_0 reprezintă frecvența centrală a benzii, $\Delta\omega$ lărgimea de bandă, iar t_0 timpul de întârziere de grup. Modulul și faza filtrului trece bandă ideal sunt prezentate în figura 5.12

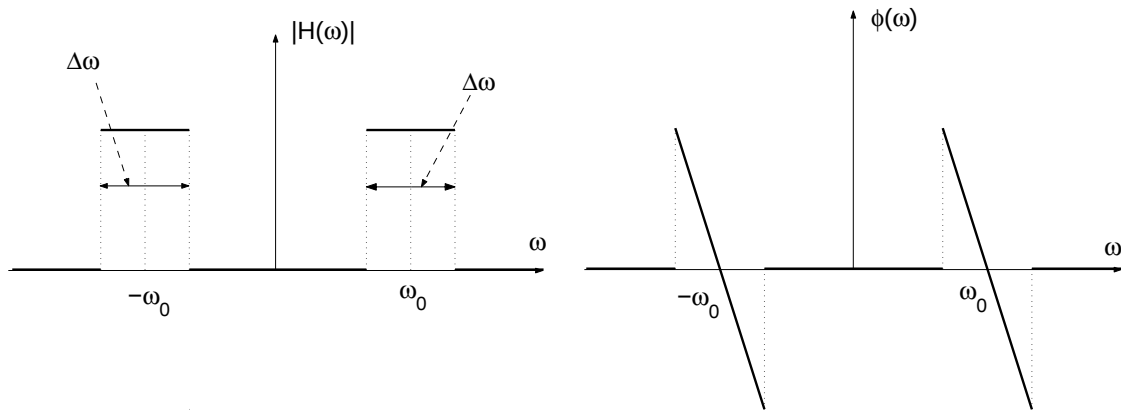


Figura 5.12: Modulul și faza unui filtru trece-bandă ideal.

Funcția pondere a FTB ideal este dată de:

$$h(t) = \mathcal{F}^{-1} \{H(\omega)\} (t) = \frac{A\Delta\omega}{\pi} \text{sinc} \left(\frac{\Delta\omega}{2}(t - t_0) \right) \cos(\omega_0 t), \quad (5.109)$$

relație ce poate fi obținută pornind de la rezultatul din ecuația (5.102) și aplicând teorema convoluției în frecvență a transformatei Fourier.

Dacă la intrarea FTB ideal se aplică semnalul aleator staționar $\xi(t)$, atunci la ieșire vom avea semnalul $\eta(t)$ având densitatea spectrală de putere și funcția de autocorelație date de:

$$q_\eta(\omega) = \begin{cases} A^2 q_\xi(\omega) & \text{dacă } |\omega| \in [\omega_0 - \frac{\Delta\omega}{2}, \omega_0 + \frac{\Delta\omega}{2}] \\ 0 & \text{în rest} \end{cases}, \quad (5.110)$$

$$R_\eta(\tau) = A^2 \left(\int_{-\omega_0 - \frac{\Delta\omega}{2}}^{-\omega_0 + \frac{\Delta\omega}{2}} q_\xi(\omega) \exp(j\omega\tau) d\omega + \int_{\omega_0 - \frac{\Delta\omega}{2}}^{\omega_0 + \frac{\Delta\omega}{2}} q_\xi(\omega) \exp(j\omega\tau) d\omega \right). \quad (5.111)$$

Dacă semnalul de intrare este zgomot alb, caracterizat de (5.80) și (5.81), atunci funcția de autocorelație a semnalului de ieșire devine:

$$R_\eta(\tau) = \frac{A^2 K \Delta\omega}{\pi} \text{sinc} \left(\frac{\Delta\omega}{2}\tau \right) \cos(\omega_0 \tau). \quad (5.112)$$

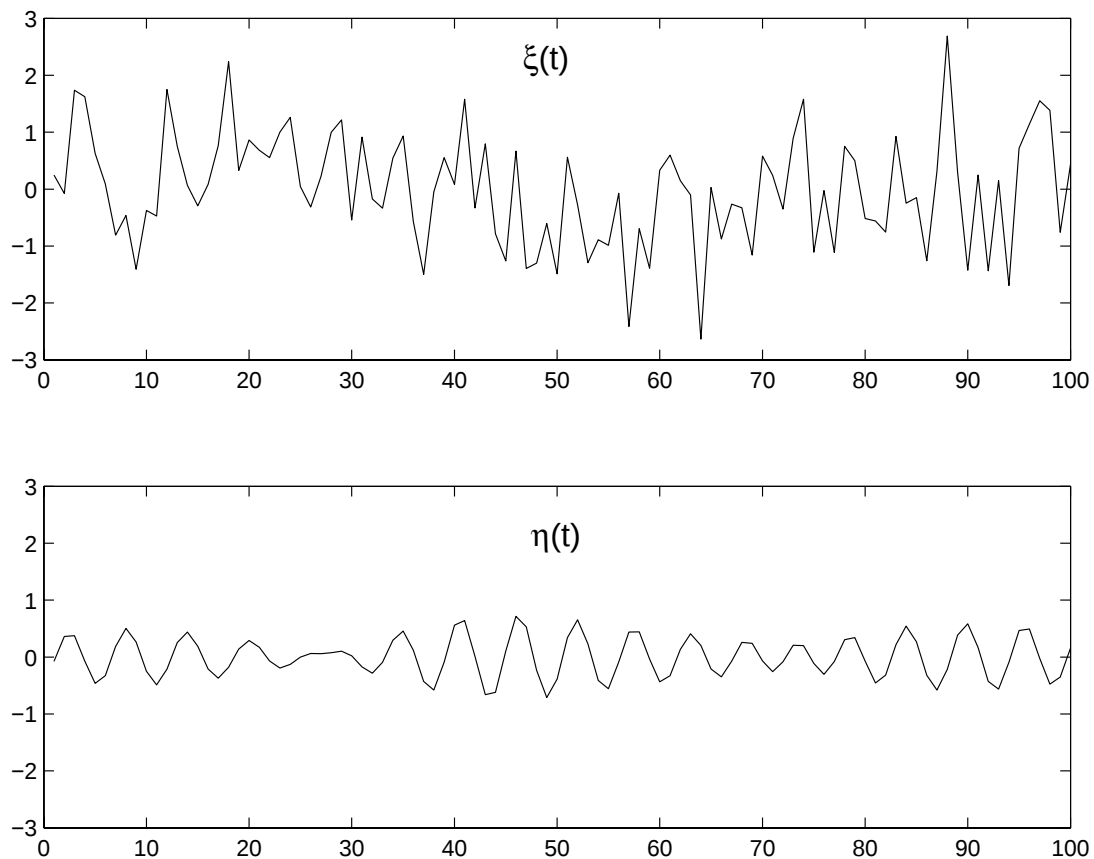


Figura 5.13: Realizare particulară a unui semnal zgomot alb $\xi(t)$ și versiunea sa filtrată cu un FTB $\eta(t)$.

Graficul funcției de autocorelație (5.112), realizat în ipoteza $\Delta\omega \ll \omega_0$ este prezentat în figura 5.14.

În figura 5.13 sunt prezentate versiunile originală și filtrată cu un FTB selectiv (cu bandă de trecere îngustă) a unei realizări particulare a unui semnal zgomot alb. Se poate ușor observa tendința filtrului de a păstra niște sinusoide de frecvențe relativ apropiate, semnalul filtrat semănând destul de bine cu o sinusoidă, spre deosebire de cel original.

În figura 5.15 este figurat norul de puncte bidimensional $[\eta(t); \eta(t + \tau)]$ pentru diverse valori ale lui τ , observându-se faptul că în semnal s-a indus o corelație pe termen destul de lung, datorată selectivității filtrului utilizat. Deci, și în cazul FTB putem trage aceleași concluzii ca și cele discutate în cazul FTJ, relative la corelația suplimentară indusă într-un semnal aleator prin operația de filtrare liniară.

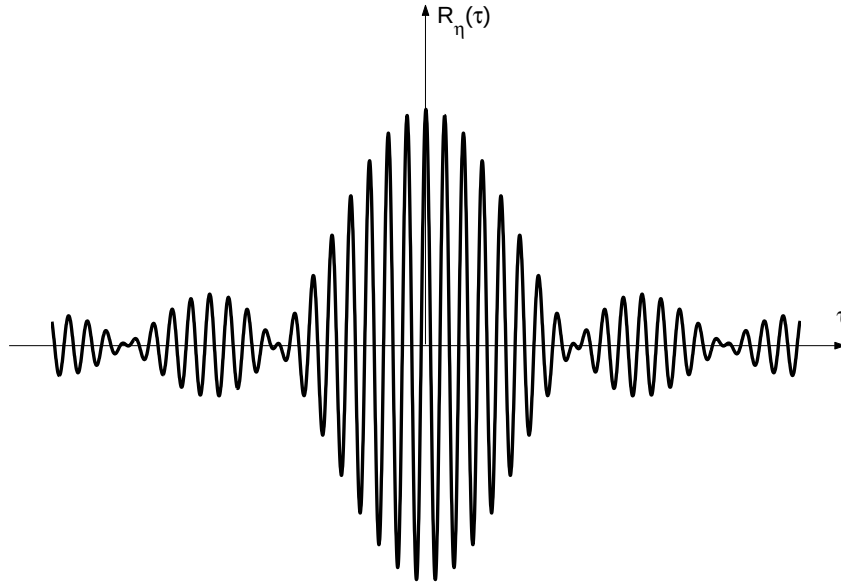


Figura 5.14: Funcția de autocorelație a semnalului de la ieșirea filtrului trece-bandă ideal având la intrare un zgomot alb.

5.12 Filtrul adaptat la semnal

Fie $s(t)$ un semnal *determinist* de durată finită. Se numește filtru adaptat la semnal filtrul liniar invariant în timp care are funcția pondere $h(t)$ dată de:

$$h(t) = Ks(-(t - t_0)), \quad (5.113)$$

unde K și t_0 sunt constante reale oarecare, singura constrângere fiind că t_0 trebuie ales astfel încât filtrul să fie cauzal, adică $h(t) = 0$ pentru $t < 0$. Un exemplu de filtru adaptat la semnal este prezentat în figura 5.16.

Funcția de transfer a filtrului adaptat la semnal este:

$$\begin{aligned} H(\omega) &= K \int_{-\infty}^{\infty} s(-(t - t_0)) \exp(-j\omega t) dt = -K \int_{\infty}^{-\infty} s(\tau) \exp(-j\omega(t_0 - \tau)) d\tau \\ &= K \exp(-j\omega t_0) \int_{-\infty}^{\infty} s(t) \exp(j\omega_0 \tau) d\tau = KS^*(\omega) \exp(-j\omega t_0), \end{aligned} \quad (5.114)$$

unde $S(\omega)$ este transformata Fourier a lui $s(t)$. În dezvoltarea de mai sus, am făcut schimbarea de variabilă $-(t - t_0) = \tau$.

În paragraful următor vom demonstra o proprietate importantă a filtrului adaptat la semnal, și anume că maximizează raportul semnal-zgomot pentru semnalul dat.

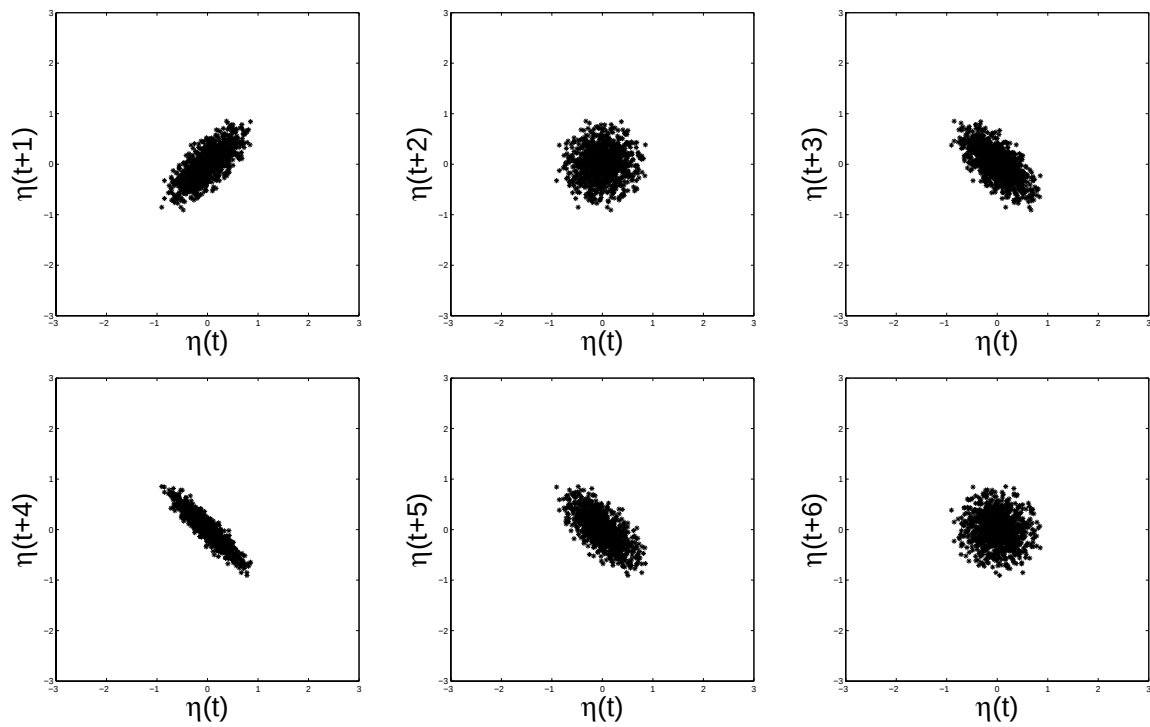


Figura 5.15: Norul de puncte $(s(t), s(t + \tau))$ pentru semnalul filtrat cu FTB $\eta(t)$ din figura 5.13 pentru $\tau = 1, \dots, 6$.

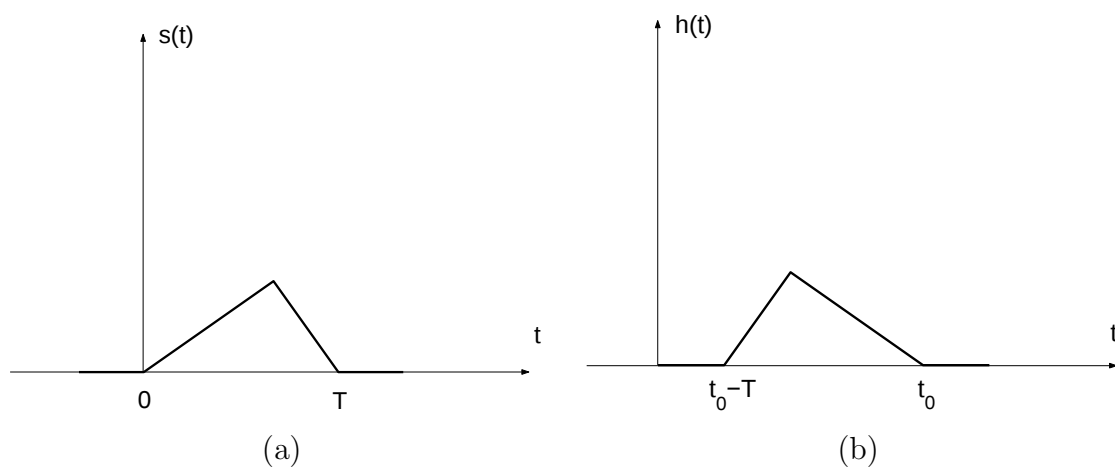


Figura 5.16: Exemplu de filtru adaptat la semnal: (a) semnalul original, (b) funcția pondere a filtrului adaptat la semnal.

5.12.1 Maximizarea raportului semnal–zgomot prin filtrare liniară

Dispunem de un semnal rezultat prin perturbarea (aditivă) cu zgomot $n(t)$ a unui semnal util $s(t)$. Aplicând acest semnal, $x(t) = s(t) + n(t)$, la intrarea unui filtru liniar invariant în timp având funcția pondere $h(t)$, obținem la ieșire un semnal pe care, datorită liniarității filtrului, îl putem interpreta ca fiind componentă utilă plus componentă remanentă a zgomotului. Așadar, putem scrie semnalul de la ieșirea filtrului sub forma $y(t) = s_o(t) + n_o(t)$, unde $s_o(t)$ este semnalul ce l-am obține la ieșire dacă la intrare am avea $s(t)$, iar $n_o(t)$ este semnalul rezultat prin trecerea prin filtru a lui $n(t)$.

Problema pe care ne-o punem în continuare este căutarea celui filtru liniar care face ca raportul semnal–zgomot după filtrare să fie maxim. În acest sens, definim raportul semnal–zgomot la un moment de timp fixat t_f ca fiind raportul dintre puterea instantanee a semnalului util la momentul respectiv de timp și puterea medie a zgomotului. Pentru semnalele de la ieșirea filtrului, aceasta se scrie:

$$\text{RSZ}_{t_f} = \frac{|s_o(t_f)|^2}{P_{n_o}}. \quad (5.115)$$

Ținând cont de faptul că

$$\begin{aligned} s_o(t_f) &= \mathcal{F}^{-1}\{S_o(\omega)\}(t_f) = \mathcal{F}^{-1}\{H(\omega)S(\omega)\}(t_f) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} H(\omega)S(\omega) \exp(j\omega t_f) d\omega, \end{aligned} \quad (5.116)$$

respectiv,

$$P_{n_o} = \frac{1}{2\pi} \int_{-\infty}^{\infty} q_{n_o}(\omega) d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} |H(\omega)|^2 q_n(\omega) d\omega, \quad (5.117)$$

relația (5.115) devine:

$$\text{RSZ}_{t_f} = \frac{1}{2\pi} \frac{\left| \int_{-\infty}^{\infty} H(\omega)S(\omega) \exp(j\omega t_f) d\omega \right|^2}{\int_{-\infty}^{\infty} |H(\omega)|^2 q_n(\omega) d\omega}. \quad (5.118)$$

În continuare, vom aplica inegalitatea Cauchy–Buniakowski–Schwartz, conform căreia, pentru orice două funcții având valori complexe $A(x)$ și $B(x)$ are loc relația:

$$\left| \int_{-\infty}^{\infty} A(x)B(x) dx \right|^2 \leq \int_{-\infty}^{\infty} |A(x)|^2 dx \int_{-\infty}^{\infty} |B(x)|^2 dx, \quad (5.119)$$

cu egalitate dacă și numai dacă

$$A(x) = kB^*(x), \quad \forall x \quad (5.120)$$

unde k este o constantă oarecare⁴. Astfel, alegem

$$A(\omega) = H(\omega)\sqrt{q_n(\omega)}, \quad (5.121)$$

respectiv,

$$B(\omega) = \frac{S(\omega) \exp(j\omega t_f)}{\sqrt{q_n(\omega)}}. \quad (5.122)$$

În relațiile (5.121) și (5.122) s-a ținut cont de faptul că densitatea spectrală de putere $q_n(\omega)$ este o funcție pozitivă. Se observă, de asemenea, că prin alegerea făcută produsul $A(\omega)B(\omega)$ reprezintă exact integrandul aflat la numărătorul expresiei (5.118). Astfel, folosind (5.119), obținem:

$$\text{RSZ}_{t_f} \leq \frac{1}{2\pi} \frac{\int_{-\infty}^{\infty} |H(\omega)\sqrt{q_n(\omega)}|^2 d\omega \int_{-\infty}^{\infty} \left| \frac{S(\omega) \exp(j\omega t_f)}{\sqrt{q_n(\omega)}} \right|^2 d\omega}{\int_{-\infty}^{\infty} |H(\omega)|^2 q_n(\omega) d\omega}, \quad (5.123)$$

de unde, prin reducerea primei integrale de la numărător cu cea de la numitor (care sunt identice), se obține:

$$\text{RSZ}_{t_f} \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| \frac{S(\omega) \exp(j\omega t_f)}{\sqrt{q_n(\omega)}} \right|^2 d\omega = \int_{-\infty}^{\infty} \frac{|S(\omega)|^2}{q_n(\omega)} d\omega \quad (5.124)$$

Se observă că integrala care limitează superior raportul semnal-zgomot nu depinde de filtrul folosit, ci numai de semnalul util și de puterea zgomotului la intrare, care sunt date. Rezultă, deci, că această integrală reprezintă raportul semnal-zgomot maxim ce se poate obține prin filtrare liniară:

$$\text{RSZ}_{t_f, \max} = \int_{-\infty}^{\infty} \frac{|S(\omega)|^2}{q_n(\omega)} d\omega. \quad (5.125)$$

Mai mult, raportul semnal-zgomot obținut își poate atinge valoarea maximă dată de (5.125) dacă în ecuația (5.124) se obține egalitate, cu alte cuvinte, când este întrunită condiția ca inegalitatea Cauchy–Buniakovski–Schwartz să se transforme în egalitate, mai precis, când relația (5.120) este îndeplinită. Ținând cont de (5.121) și (5.122), aceasta devine:

$$H(\omega)\sqrt{q_n(\omega)} = k \left(\frac{S(\omega) \exp(j\omega t_f)}{\sqrt{q_n(\omega)}} \right)^*, \quad (5.126)$$

⁴Relațiile (5.119) și (5.120) nu reprezintă altceva decât exprimarea în formă continuă, pentru valori complexe, a relației cunoscute din liceu, conform căreia, pentru orice numere reale $a_1, \dots, a_N$, respectiv $b_1, \dots, b_N$, are loc inegalitatea $\sum_{i=1}^N (a_i b_i)^2 \leq \left(\sum_{i=1}^N a_i^2 \right) \left(\sum_{i=1}^N b_i^2 \right)$, cu egalitate dacă și numai dacă $\frac{a_1}{b_1} = \frac{a_2}{b_2} = \dots = \frac{a_N}{b_N}$.

ceea ce, ținând cont că $q_n(\omega)$ e reală, se scrie:

$$H(\omega) = k \frac{S^*(\omega) \exp(-j\omega t_f)}{q_n(\omega)}. \quad (5.127)$$

Presupunând mai departe că zgomotul care perturbă semnalul este zgomot alb, și deci, q_n este de forma (5.80), rezultă că funcția de transfer a filtrului care maximizează raportul semnal-zgomot la ieșire devine:

$$H(\omega) = k_1 S^*(\omega) \exp(-j\omega t_f). \quad (5.128)$$

Readucându-ne aminte de relația (5.114), putem concluziona că *filtrul care maximizează raportul semnal-zgomot pentru un semnal util dat este chiar filtrul adaptat la semnalul util*.

Acest rezultat are câteva aplicații practice importante, dintre care menționăm aici pe scurt numai detecția semnalelor, unde problema care se pune este de a decide asupra transmiterii unui semnal dintr-o mulțime finită de semnale cunoscute în prezența zgomotului. Mai multe detalii despre acest subiect vor fi prezentate în capitolul 6.

Capitolul 6

Detecția semnalelor

6.1 Introducere

Modelul unui lanț de transmisiune cu detecție pe care îl vom lua în considerare în acest capitol este prezentat în figura 6.1. Sursa **S** este o sursă discretă de informație, care emite simboluri S_i dintr-o mulțime cu M simboluri $\{S_0, S_1, \dots, S_{M-1}\}$ cu probabilități $P(S_i) \stackrel{\text{not}}{=} P_i$ presupuse cunoscute. Pentru fiecare simbol S_i , generatorul de semnal **G** generează un semnal $s_i(t)$ pe durata $t \in [0, T]$, care este transmis pe canal. În timpul transmisiunii, semnalul util este afectat de zgomot $n(t)$, care este un semnal aleator perturbator de natură aditivă, astfel încât semnalul recepționat la ieșirea din canal se poate scrie $r(t) = s_i(t) + n(t)$ pentru $t \in [0, T]$. Problema detecției, respectiv a implementării blocului de decizie **D**, este ca pe baza semnalului recepționat $r(t)$ pe o durată $[0, T]$ să se decidă asupra simbolului emis de sursă. Ieșirea blocului de decizie către utilizatorul final **U** este o decizie D_i , care semnifică faptul că se *decide* că s-a transmis simbolul S_i (ceea ce, bineînțeles, poate corespunde sau nu realității).

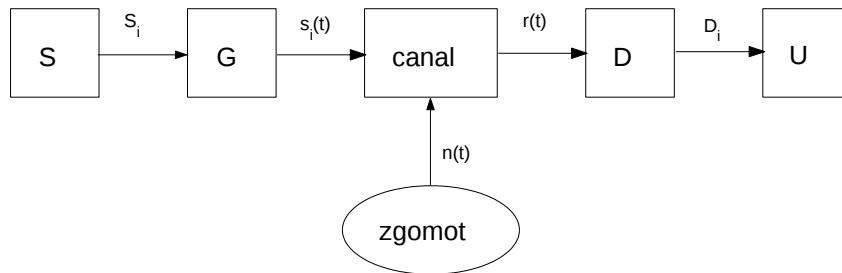


Figura 6.1: Schema generală a unui lanț de transmisiune cu detecție.

În figura 6.2 este prezentat un exemplu pentru semnalele de la intrarea și ieșirea din canal pentru un caz particular: sursă binară ($M = 2$), semnale constante $s_0(t) = 0$, $s_1(t) = A$ pentru $t \in [0, T]$ alocate celor două simboluri. Se observă, fie și numai din punct de vedere calitativ, că problema deciziei asupra simbolului emis de sursă pornind de la semnalul recuperat la ieșirea canalului nu este trivială, în urma perturbării cu zgomot,

acesta semănând destul de puțin cu versiunea sa “curată”.

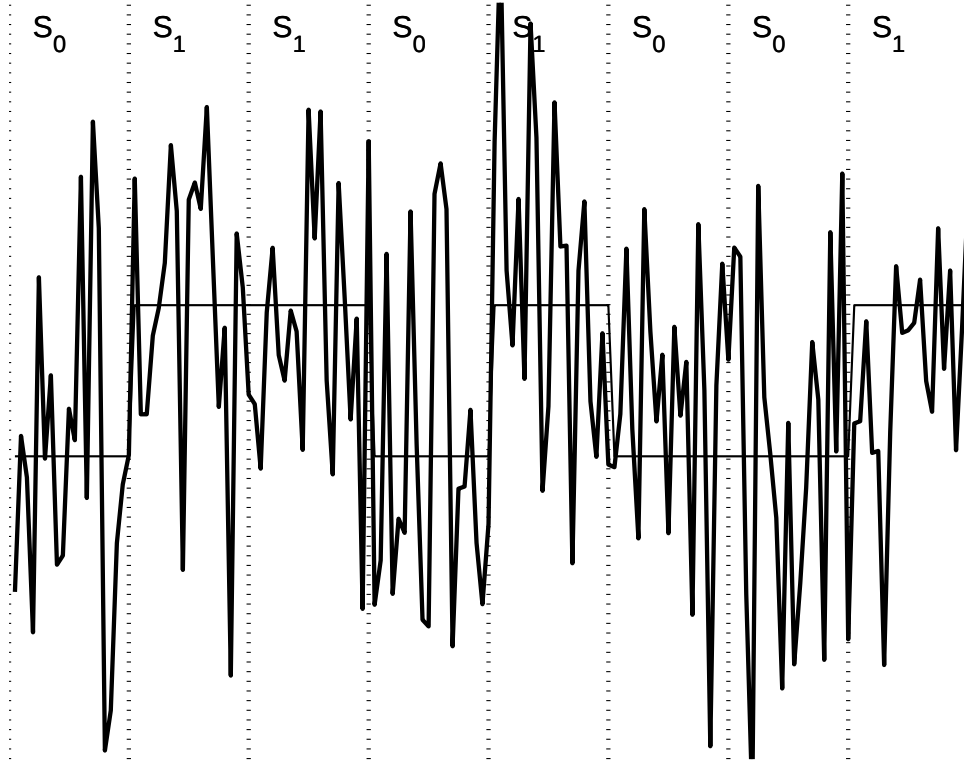


Figura 6.2: Exemplu de transmisiune a unui semnal binar printr-un canal cu zgomot. S-a considerat $s_0(t) = 0$, $s_1(t) = A$. Cu linie subțire este prezentat semnalul corespunzător secvenței binare $S_0 S_1 S_1 S_0 S_1 S_0 S_0 S_1$ la intrarea pe canal, iar cu linie îngroșată semnalul la ieșirea din canal.

Semnalele $s_i(t)$ emise de generatorul $\mathbf{G}$ pot fi deterministe, cunoscute atât la emisie, cât și la recepție (caz în care spunem că ne aflăm într-o ipoteză *simplă*) sau pot fi la rândul lor aleatoare, caz în care spunem că ne aflăm într-o ipoteză *compusă*. În ipoteza compusă, ceea ce diferă între semnalele $s_i(t)$ pentru $i = 0, \dots, M - 1$ sunt parametrii statistici ai acestora.

La recepție, decizia poate fi luată pe baza semnalului $r(t)$ observat în mod continuu, pentru $t \in [0, T]$, sau acesta poate fi discretizat (spunem că spațiul observațiilor este discret). În acest caz, se alege un număr de N momente de timp în intervalul $[0, T]$:

$$0 \leq t_1 < t_2 < \dots < t_N \leq T,$$

și se rețin numai valorile semnalului recepționat la acele momente: $r_i \stackrel{\text{not}}{=} r(t_i)$ pentru $i = 1, \dots, N$. În acest caz, decizia se va lua pe baza vectorului aleator N -dimensional având drept componente cele N valori ale semnalului recepționat:

$$\mathbf{r} = (r_1, r_2, \dots, r_N).$$

În continuarea acestui capitol, vom discuta despre regula de decizie Bayes în ipoteza binară simplă, caz în care alfabetul sursei este binar: $\{S_0, S_1\}$, iar semnalele $s_i(t)$ cu $i = 0, 1$ sunt deterministe.

6.2 Criteriul de decizie Bayes: cazul observațiilor discrete

În acest paragraf ne vom ocupa de deducerea criteriului de decizie Bayes considerând spațiul observațiilor discret. Pentru fiecare simbol emis de sursă decizia se va lua, deci, pe baza vectorului recepționat $\mathbf{r} = (r_1, r_2, \dots, r_N)$. Evident, datorită zgomotului, care este un proces aleator, observația $\mathbf{r}$ va diferi de la simbol la simbol, chiar în ipoteza transmiterii aceluiași simbol S_i de către sursă, ceea ce ne conduce la a-l considera pe $\mathbf{r}$ ca fiind o realizare particulară a unui vector aleator N -dimensional, pe care-l vom nota $\mathbf{R} = (R_1, R_2, \dots, R_N)$:

$$\mathbf{r} \equiv \mathbf{R}^{(k)}. \quad (6.1)$$

Deducerea unui criteriu de decizie în aceste condiții constă în calculul unei hipersuprafețe de decizie în $\mathbb{R}^N$, care să-l împartă pe acesta în două zone de decizie Δ_0 și Δ_1 , cu $\Delta_0 \cap \Delta_1 = \emptyset$, respectiv $\Delta_0 \cup \Delta_1 = \mathbb{R}^N$. Astfel, dacă vectorul recepționat $\mathbf{r} \in \Delta_0$ se decide D_0 , iar dacă $\mathbf{r} \in \Delta_1$ se decide D_1 . Canalul de transmisiune astfel obținut este un canal binar¹.

Pentru ilustrarea geometrică a problemei, în scopul înțelegerii acesteia, să presupunem că avem următoarele date: $s_0(t) = 0$, $s_1(t) = A$, pentru $t \in [0, T]$ și că la recepție luăm decizia pe baza a $N = 2$ eșantioane, deci $\mathbf{r} = (r_1, r_2)$. În absența zgomotului, avem $r(t) = s_o(t) = 0$, dacă sursa emite S_0 și, deci $\mathbf{R}|S_0 = [0, 0]$, respectiv $r(t) = s_1(t) = A$ pentru simbolul S_1 ceea ce conduce la $\mathbf{R}|S_1 = [A, A]$. Situația este ilustrată în figura 6.3.(a). În prezența zgomotului, avem $\mathbf{R}|S_0 = [n_1, n_2]$, respectiv $\mathbf{R}|S_1 = [A+n_1, A+n_2]$ unde n_1 și n_2 reprezintă valorile zgomotului la cele două momente alese pentru eșantionarea semnalului $r(t)$. Norii de puncte recepționate ($\mathbf{R}^{(i)}|S_0$, respectiv $\mathbf{R}^{(i)}|S_1$) obținuți la transmisiunea unui număr mare de simboluri în acest caz particular sunt ilustrați în figura 6.3.(b). S-a considerat în acest exemplu că cele două eșantioane de zgomot n_1 și n_2 sunt independente, de unde rezultă structura haotică a celor două aglomerări de puncte. Dreapta punctată din figura 6.3.(b)² pare cea mai naturală alegere (dintr-un punct de vedere strict calitativ) pentru împărțirea lui $\mathbb{R}^2$ în cele două zone de decizie Δ_0 (semiplanul de sub dreaptă) și Δ_1 (semiplanul de deasupra dreptei)³. Se observă, de asemenea, că există

¹S-ar mai putea lua în considerare și împărțirea lui $\mathbb{R}^N$ în trei zone de decizie: Δ_0 , Δ_1 , respectiv Δ_{anulare} . În acest caz, dacă $\mathbf{r} \in \Delta_{\text{anulare}}$, nu se ia nici o decizie asupra simbolului transmis de sursă, considerându-se că o decizie ar fi prea riscantă. Canalul de transmisiune astfel obținut devine un canal cu alfabet binar la intrare și alfabet ternar la ieșire, care, se numește, după caz, fie canal binar cu anulări, fie canal binar cu erori și anulări.

²În $\mathbb{R}^2$, o hipersuprafață este o curbă, în $\mathbb{R}^3$ o suprafață, etc.

³Dacă s-ar considera pentru acest caz și zona de decizie Δ_{anulare} , aceasta ar trebui să fie o bandă delimitată de două drepte paralele cu dreapta punctată din figură, situate de-o parte și de cealaltă a

puncte clasificate greșit (puncte “ $\diamond$ ” în Δ_1 , respectiv puncte “ $\circ$ ” în Δ_0 , acestea corespunzând unor decizii greșite.

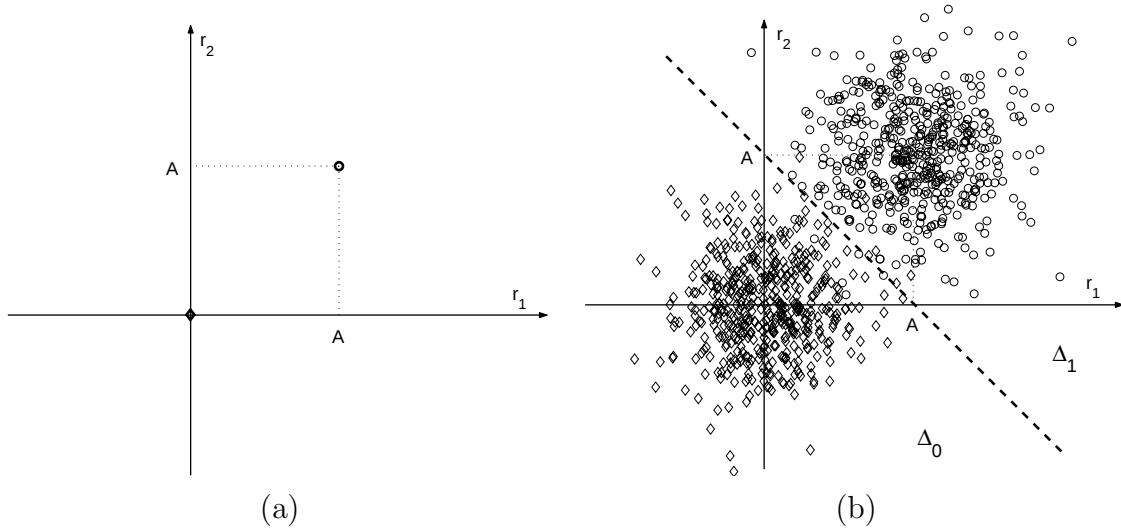


Figura 6.3: Exemplificare 2D a procesului de decizie. Cu $\diamond$ sunt figurate punctele $\mathbf{R}^{(i)}|S_0$, iar cu $\circ$ punctele $\mathbf{R}^{(i)}|S_1$, în două cazuri: (a) în absența zgomotului și (b) în prezența zgomotului.

Împărțirea lui $\mathbb{R}^N$ în cele două zone de decizie trebuie făcută în sensul minimizării unui criteriu de eroare. În acest sens, se presupune că deciziile sunt legate de niște costuri presupuse cunoscute (impuse de aplicație) C_{ij} cu $i, j = 0, 1$, care sunt numere pozitive ($C_{ij} \geq 0$) care reprezintă costul deciziei D_j când sursa emite S_i :

$$C_{ij} = \text{Cost}(S_i \cap D_j), \quad i, j = 0, 1.$$

Cu cât este de dorit ca un eveniment $S_i \cap D_j$ să se întâmple mai rar, cu atât costul aferent C_{ij} este mai mare⁴.

Regula de decizie Bayes se ia în sensul minimizării costului mediu al deciziei. Odată definite costurile aferente fiecărui eveniment $S_i \cap D_j$ ce poate surveni la transmisia unui simbol pe canal, se definește valoarea medie a costului de decizie $\bar{C}$:

$$\begin{aligned} \bar{C} &= C_{00}P(D_0 \cap S_0) + C_{01}P(D_1 \cap S_0) + C_{10}P(D_0 \cap S_1) + C_{11}P(D_1 \cap S_1) \\ &= C_{00}P(D_0|S_0)P_0 + C_{01}P(D_1|S_0)P_0 + C_{10}P(D_0|S_1)P_1 + C_{11}P(D_1|S_1)P_1, \end{aligned} \quad (6.2)$$

acesteia.

⁴Pentru majoritatea transmisiunilor digitale, evenimentele $S_0 \cap D_1$ (respectiv, transmiterea eronată a unui “0”) și $S_1 \cap D_0$ (respectiv, transmiterea eronată a unui “1”) sunt la fel de nedorite, caz în care trebuie ales $C_{01} = C_{10}$. De asemenea, alegerea naturală pentru costurile evenimentelor “transmisie corectă” este $C_{00} = C_{11} = 0$. Există însă cazuri în care importanța deciziilor eronate asupra simbolurilor S_0 și S_1 să difere; atunci $C_{01} \neq C_{10}$ (un exemplu tipic este detecția țintelor aeriene cu un sistem radar).

iar criteriul de decizie Bayes urmărește, după cum am enunțat anterior, minimizarea acestui cost mediu $\bar{C}$. În calculul de la (6.2), am folosit formula probabilității condiționate (2.4).

În continuare, pentru simplificarea formalismului matematic, vom folosi următoarele convenții de notație simplificate:

$$\begin{aligned} w_{\mathbf{R}}(\mathbf{r}) &\stackrel{\text{not}}{=} w_{R_1 R_2 \dots R_N}(r_1, \dots, r_N) \\ d\mathbf{r} &\stackrel{\text{not}}{=} dr_1 dr_2 \dots dr_N \\ \int &\stackrel{\text{not}}{=} \int \dots \int. \end{aligned} \quad (6.3)$$

Astfel, în aceste condiții, putem scrie:

$$P(D_0|S_0) = P((\mathbf{R}|S_0) \in \Delta_0) = \int_{\Delta_0} w_{\mathbf{R}|S_0}(\mathbf{r}) d\mathbf{r}, \quad (6.4)$$

unde $w_{\mathbf{R}|S_0}$ reprezintă densitatea de probabilitate de ordinul N a variabilelor aleatoare $R_1, \dots, R_N$ în ipoteza emiterii simbolului S_0 (cu alte cuvinte, revenind la exemplul din figura 6.3, densitatea de probabilitate a punctelor figurate cu $\diamond$). Relația (6.4) nu reprezintă altceva decât extinderea la cazul a N variabile aleatoare a proprietății 2 a densității de probabilitate de ordinul doi (prezentată la pagina 34). În mod absolut similar, și celelalte probabilități $P(D_j|S_i)$ pot fi scrise:

$$P(D_j|S_i) = P((\mathbf{R}|S_i) \in \Delta_j) = \int_{\Delta_j} w_{\mathbf{R}|S_i}(\mathbf{r}) d\mathbf{r}, \quad i, j = 0, 1, \quad (6.5)$$

ceea ce conduce la următoarea relație pentru costul mediu:

$$\bar{C} = \sum_{i=0}^1 \sum_{j=0}^1 C_{ij} P_i \int_{\Delta_j} w_{\mathbf{R}|S_i}(\mathbf{r}) d\mathbf{r}. \quad (6.6)$$

În continuare, ținând cont că:

$$\int_{\Delta_0} w_{\mathbf{R}|S_0}(\mathbf{r}) d\mathbf{r} + \int_{\Delta_1} w_{\mathbf{R}|S_0}(\mathbf{r}) d\mathbf{r} = \int_{\mathbb{R}^N} w_{\mathbf{R}|S_0}(\mathbf{r}) d\mathbf{r} = P((\mathbf{R}|S_0) \in \mathbb{R}^N) = 1, \quad (6.7)$$

putem scrie în relația (6.6) toate integralele pe domeniul Δ_1 ca niște integrale pe Δ_0 , după:

$$\int_{\Delta_1} w_{\mathbf{R}|S_i}(\mathbf{r}) d\mathbf{r} = 1 - \int_{\Delta_0} w_{\mathbf{R}|S_i}(\mathbf{r}) d\mathbf{r}, \quad i = 0, 1. \quad (6.8)$$

Astfel, valoarea costului mediu devine:

$$\begin{aligned} \overline{C} = C_{00}P_0 \int_{\Delta_0} w_{\mathbf{R}|S_0}(\mathbf{r})d\mathbf{r} + C_{10}P_1 \int_{\Delta_0} w_{\mathbf{R}|S_1}(\mathbf{r})d\mathbf{r} + \\ + C_{01}P_0 \left(1 - \int_{\Delta_0} w_{\mathbf{R}|S_0}(\mathbf{r})d\mathbf{r} \right) + C_{11}P_1 \left(1 - \int_{\Delta_0} w_{\mathbf{R}|S_1}(\mathbf{r})d\mathbf{r} \right), \quad (6.9) \end{aligned}$$

ceea ce poate fi rescris mai compact drept:

$$\overline{C} = C_{01}P_0 + C_{11}P_1 + \int_{\Delta_0} (P_1(C_{10} - C_{11})w_{\mathbf{R}|S_1}(\mathbf{r}) - P_0(C_{01} - C_{00})w_{\mathbf{R}|S_0}(\mathbf{r}))d\mathbf{r}. \quad (6.10)$$

Din punct de vedere matematic, problema minimizării costului mediu $\overline{C}$ din (6.10) este una de minimizare a ariei unui subgrafic în funcție de domeniul ales. Soluția poate fi ușor văzută dacă facem o analogie unidimensională. În figura 6.4 este prezentat graficul unei funcții oarecare $f(x)$. Se observă că alegând domeniul D_1 ca în figură, respectiv $D_1 = \{x \in \mathbb{R} | f(x) < 0\}$ obținem $\int_{D_1} f(x)dx \rightarrow \min$.

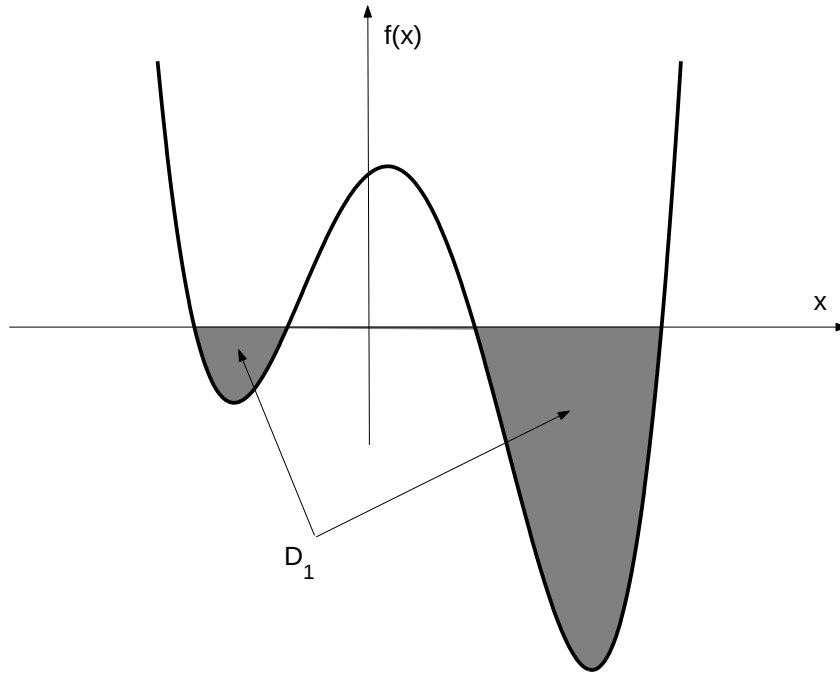


Figura 6.4: Exemplificarea 1D a alegerii domeniului care minimizează aria de sub graficul unei funcții.

Într-adevăr, orice alt domeniu D_2 am adăuga la D_1 , avem $\int_{D_1 \cup D_2} f(x)dx > \int_{D_1} f(x)dx$, întrucât $\int_{D_2} f(x)dx > 0$.

Revenind la problema alegerii intervalului Δ_0 care minimizează costul mediu $\overline{C}$ din (6.10), rezultă că putem scrie, prin analogie cu situația ilustrată mai sus:

$$\overline{C} = \overline{C}_{\min} \Leftrightarrow \Delta_0 = \left\{ \mathbf{r} \in \mathbb{R}^N \mid P_1(C_{10} - C_{11})w_{\mathbf{r}|S_1}(\mathbf{r}) - P_0(C_{01} - C_{00})w_{\mathbf{r}|S_0}(\mathbf{r}) < 0 \right\}. \quad (6.11)$$

Rezultă, deci, că domeniul Δ_1 care minimizează costul mediu este dat de:

$$\Delta_1 = \left\{ \mathbf{r} \in \mathbb{R}^N \mid P_1(C_{10} - C_{11})w_{\mathbf{r}|S_1}(\mathbf{r}) - P_0(C_{01} - C_{00})w_{\mathbf{r}|S_0}(\mathbf{r}) > 0 \right\}. \quad (6.12)$$

Având în vedere că $\mathbf{r} \in \Delta_0 \Rightarrow D_0$, respectiv $\mathbf{r} \in \Delta_1 \Rightarrow D_1$, relațiile (6.11) și (6.12) pot fi scrise mai compact astfel:

$$P_1(C_{10} - C_{11})w_{\mathbf{r}|S_1}(\mathbf{r}) \underset{D_0}{\overset{D_1}{\gtrless}} P_0(C_{01} - C_{00})w_{\mathbf{r}|S_0}(\mathbf{r}), \quad (6.13)$$

ceea ce, prin rearanjarea termenilor, conduce la forma finală a criteriului de decizie Bayes:

$$\frac{w_{\mathbf{r}|S_1}(\mathbf{r})}{w_{\mathbf{r}|S_0}(\mathbf{r})} \underset{D_0}{\overset{D_1}{\gtrless}} \frac{P_0(C_{01} - C_{00})}{P_1(C_{10} - C_{11})}. \quad (6.14)$$

Termenul din stânga se notează cu $\Lambda(\mathbf{r})$ și se numește raport de plauzibilitate:

$$\Lambda(\mathbf{r}) \stackrel{\text{not}}{=} \frac{w_{\mathbf{r}|S_1}(\mathbf{r})}{w_{\mathbf{r}|S_0}(\mathbf{r})} \quad (6.15)$$

în timp ce termenul din dreapta (care este constant) se notează cu K și se numește pragul testului:

$$K \stackrel{\text{not}}{=} \frac{P_0(C_{01} - C_{00})}{P_1(C_{10} - C_{11})}. \quad (6.16)$$

Conform relației (6.14), rezultă schema bloc a blocului de decizie **D** din figura 6.5. Prima operație este extragerea eșantioanelor r_i cu $i = 1, \dots, N$ din semnalul continuu $r(t)$, pentru $t \in [0, T]$, apoi se calculează raportul de plauzibilitate conform relației 6.15, decizia finală fiind dată de rezultatul comparației acestuia cu pragul testului dat de 6.16.

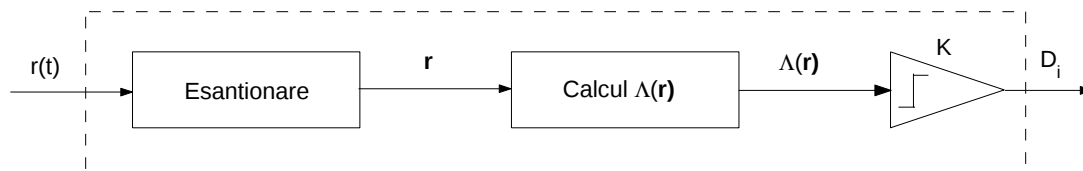


Figura 6.5: Schema bloc a blocului de detecție **D** din figura 6.1 conform regulii de decizie Bayes.

Să mai menționăm doar faptul că, uneori, se preferă exprimarea regulii de decizie Bayes în formă logaritmică, aceasta conducând la o expresie matematic mai simplă. Având în

vedere că funcția logaritm natural este strict crescătoare, se poate rescrie relația (6.14) sub forma:

$$\ln \Lambda(\mathbf{r}) \underset{D_0}{\overset{D_1}{\geq}} \ln K, \quad (6.17)$$

ce reprezintă exprimarea criteriului Bayes în formă logaritmică.

6.2.1 Statistica suficientă

Se poate face întotdeauna o schimbare de coordonate în $\mathbb{R}^N$ astfel încât toată informația necesară deciziei să fie conținută într-o singură coordonată din spațiul transformat. Acea coordonată se numește *statistică suficientă*.

Astfel, dacă se face schimbarea de coordonate $[R_1, \dots, R_N] \Leftrightarrow \underbrace{[L_1, \dots, L_N]}_{\mathbf{L}}$, astfel încât,

$$w_{L_2, \dots, L_N | S_0}(l_2, \dots, l_N) = w_{L_2, \dots, L_N | S_1}(l_2, \dots, l_N), \quad (6.18)$$

ceea ce, altfel spus, semnifică faptul că coordonatele $L_2, \dots, L_N$ nu conțin informație relevantă la decizie, atunci coordonata L_1 este statistica suficientă, iar regula de decizie Bayes se scrie:

$$\frac{w_{\mathbf{L} | S_1}(l_1, \dots, l_N)}{w_{\mathbf{L} | S_0}(l_1, \dots, l_N)} = \frac{w_{L_1 | S_1}(l_1)}{w_{L_1 | S_0}(l_1)} \underset{D_0}{\overset{D_1}{\geq}} \frac{P_0(C_{01} - C_{00})}{P_1(C_{10} - C_{11})}. \quad (6.19)$$

6.2.2 Criteriul Bayes pentru zgomot alb gaussian

În acest paragraf, vom particulariza regula de decizie Bayes pentru una din ipotezele statistice cel mai des folosite pentru zgomotul de pe canal, și anume vom presupune că $n(t)$ este un semnal aleator staționar, zgomot alb, având o distribuție de ordinul unu gaussiană: $n(t) : \mathcal{N}(0, \sigma_n)$ pentru $\forall t \in [0, T]$.

Pentru a păstra generalitatea, nu vom particulariza cele două semnale $s_0(t)$ și $s_1(t)$, mărginindu-ne la a le presupune cunoscute.

În aceste ipoteze, densitățile de probabilitate $w_{\mathbf{R} | S_i}(\mathbf{r})$ ce intervin în regula de decizie se calculează după cum urmează.

În ipoteza transmiterii lui S_0 avem $r(t) = s_0(t) + n(t)$ pentru $t \in [0, T]$, ceea ce se poate scrie în formă eșantionată:

$$R_i | S_0 = s_{0i} + n_i \quad i = 1, \dots, N, \quad (6.20)$$

unde $s_{0i} \stackrel{\text{not}}{=} s_0(t_i)$ reprezintă valorile *cunoscute* ale semnalului s_0 la momentele de timp alese pentru eșantionarea lui $r(t)$, iar $n_i \stackrel{\text{not}}{=} n(t_i)$ reprezintă valorile *aleatoare* ale zgomotului la aceleași momente.

Cu aceste precizări, din relația (6.20) rezultă că variabila aleatoare $R_i | S_0$, a cărei realizare particulară este r_i , rezultă prin sumarea unei variabile aleatoare gaussiene n_i peste

o constantă s_{0i} ⁵. Aplicând rezultatul (3.32), rezultă că și $R_i|S_0$ are o distribuție gaussiană, cu media dată de suma dintre media inițială și constanta adunată: $R_i|S_0: \mathcal{N}(s_{0i}, \sigma_n)$, ceea ce este echivalent cu:

$$w_{R_i|S_0}(r_i) = \frac{1}{\sigma_n \sqrt{2\pi}} \exp \left[-\frac{(r_i - s_{0i})^2}{2\sigma_n^2} \right] \quad i = 1, \dots, N. \quad (6.21)$$

Avem, deci, distribuțiile de ordinul unu ale tuturor componentelor vectorului $\mathbf{R}|S_0$. Pentru a putea aplica, însă, criteriul Bayes, avem nevoie de distribuția de ordinul N a tuturor componentelor. Or, după cum știm din capitolul 4, cunoștințele despre distribuțiile de ordin inferior nu oferă suficientă informație pentru calculul distribuțiilor de ordin mai mare, decât într-o singură ipoteză: cea de independență între variabilele aleatoare. Cum $n(t)$ este modelat ca fiind zgomot alb, ceea ce, conform discuției din paragraful 5.10 înseamnă ca variabilele $n(t_i)$ și $n(t_j)$ sunt decorelate pentru $t_i \neq t_j$, și cum, în plus amândouă sunt gaussiene, putem presupune că ele sunt și independente statistic⁶. Rezultă că și variabilele aleatoare $R_i|S_0$ și $R_j|S_0$ sunt independente pentru $i \neq j$, întrucât ele rezultă prin sumarea unor constante peste n_i , respectiv n_j . Putem, deci, scrie:

$$w_{\mathbf{R}|S_0}(\mathbf{r}) = \prod_{i=1}^N w_{R_i|S_0}(r_i) = \left(\frac{1}{\sigma_n \sqrt{2\pi}} \right)^N \exp \left[-\frac{1}{2\sigma_n^2} \sum_{i=1}^N (r_i - s_{0i})^2 \right]. \quad (6.22)$$

Raționamentul se poate relua în mod absolut similar și pentru $w_{\mathbf{R}|S_1}(\mathbf{r})$:

$$w_{\mathbf{R}|S_1}(\mathbf{r}) = \prod_{i=1}^N w_{R_i|S_1}(r_i) = \left(\frac{1}{\sigma_n \sqrt{2\pi}} \right)^N \exp \left[-\frac{1}{2\sigma_n^2} \sum_{i=1}^N (r_i - s_{1i})^2 \right]. \quad (6.23)$$

Cu densitățile de probabilitate deduse, rezultă că raportul de plauzibilitate se scrie:

$$\begin{aligned} \Lambda(\mathbf{r}) &= \frac{w_{\mathbf{R}|S_1}(\mathbf{r})}{w_{\mathbf{R}|S_0}(\mathbf{r})} = \frac{\left(\frac{1}{\sigma_n \sqrt{2\pi}} \right)^N \exp \left[-\frac{1}{2\sigma_n^2} \sum_{i=1}^N (r_i - s_{1i})^2 \right]}{\left(\frac{1}{\sigma_n \sqrt{2\pi}} \right)^N \exp \left[-\frac{1}{2\sigma_n^2} \sum_{i=1}^N (r_i - s_{0i})^2 \right]} \\ &= \exp \left[-\frac{1}{2\sigma_n^2} \sum_{i=1}^N ((r_i - s_{1i})^2 - (r_i - s_{0i})^2) \right] \\ &= \exp \left[-\frac{1}{2\sigma_n^2} \sum_{i=1}^N (r_i^2 - 2r_i s_{1i} + s_{1i}^2 - r_i^2 + 2r_i s_{0i} - s_{0i}^2) \right] \\ &= \exp \left[\frac{1}{\sigma_n^2} \sum_{i=1}^N r_i (s_{1i} - s_{0i}) - \frac{1}{2\sigma_n^2} \sum_{i=1}^N (s_{1i}^2 - s_{0i}^2) \right]. \end{aligned} \quad (6.24)$$

⁵În ipoteza transiterii lui S_0 , valorile s_{0i} sunt constante din punct de vedere *statistic* și nu temporal. Altfel spus, în expresia celui de-al i -lea eșantion al semnalului recepționat în ipoteza S_0 va apărea întotdeauna valoarea s_{0i} .

⁶Pentru ca $n(t_i)$ și $n(t_j)$ să fie independente, ar trebui ca și modelul statistic de ordinul doi al zgomotului să fie tot gaussian. În practică, cum nu există semnal pur aleator, nu se poate presupune decorelarea sau independența între $n(t_i)$ și $n(t_j)$ pentru $t_j \rightarrow t_i$. Se poate alege, însă, ecartul temporal minim între două momente succesive t_{i-1} și t_i suficient de mare ca corelația din semnal să se “stingă”; astfel, ipoteza independenței între valorile $n(t_i)$ devine verosimilă.

Se observă că pentru cazul de față este de preferat exprimarea criteriului Bayes în forma sa logaritmică. Astfel, relația (6.17) devine:

$$\frac{1}{\sigma_n^2} \sum_{i=1}^N r_i(s_{1i} - s_{0i}) - \frac{1}{2\sigma_n^2} \sum_{i=1}^N (s_{1i}^2 - s_{0i}^2) \underset{D_0}{\overset{D_1}{\geq}} \ln K, \quad (6.25)$$

care, prin prelucrări elementare, poate fi scrisă sub formă finală ca:

$$\sum_{i=1}^N r_i(s_{1i} - s_{0i}) \underset{D_0}{\overset{D_1}{\geq}} \sigma_n^2 \ln K + \frac{1}{2} \sum_{i=1}^N (s_{1i}^2 - s_{0i}^2), \quad (6.26)$$

Observație. Termenul din stânga al relației (6.26), $l_1 \stackrel{\text{not}}{=} \sum_{i=1}^N r_i(s_{1i} - s_{0i})$ reprezintă statistica suficientă în acest caz, respectiv aceea coordonată unică ce conține toată informația necesară deciziei.

Exemplu. Să particularizăm relația (6.26) pentru următoarele semnale: $s_0(t) = 0$, $s_1(t) = A$ pentru $t \in [0, T]$. Astfel, avem $s_{0i} = 0$ și $s_{1i} = A$ pentru $\forall i = 1, \dots, N$. Să mai presupunem că $P_0 = P_1 = \frac{1}{2}$ (simboluri echiprobabile), $C_{00} = C_{11}$ (costuri egale pentru deciziile corecte) și $C_{01} = C_{10}$ (costuri egale pentru deciziile greșite), ceea ce conduce la $K = 1$, respectiv $\ln K = 0$. În aceste condiții, relația (6.26) se scrie:

$$\sum_{i=1}^N r_i(A - 0) \underset{D_0}{\overset{D_1}{\geq}} \frac{1}{2} \sum_{i=1}^N (A^2 - 0), \quad (6.27)$$

ceea ce revine la:

$$\frac{1}{N} \sum_{i=1}^N r_i \underset{D_0}{\overset{D_1}{\geq}} \frac{A}{2}. \quad (6.28)$$

Revenind la exemplul prezentat în figura 6.3.(b), pentru $N = 2$, relația (6.28) se scrie $r_1 + r_2 \underset{D_0}{\overset{D_1}{\geq}} A$, ceea ce corespunde soluției găsite pe baza argumentelor calitative (ecuația dreptei punctate din figură este $r_1 + r_2 = A$).

În acest punct este utilă discuția influenței costurilor asupra deciziei. Să presupunem că alegem $C_{01} > C_{10}$, ceea ce este un mod de a “comunica” sistemului de decizie că evenimentul $S_0 \cap D_1$ este mai de nedorit decât $S_1 \cap D_0$. În aceste condiții, avem $K > 1$, respectiv $\ln K > 0$, iar regula de decizie revine $r_1 + r_2 \underset{D_0}{\overset{D_1}{\geq}} A + \alpha$ cu $\alpha = \sigma_n^2 \ln K > 0$. Altfel spus, dreapta care separă cele două zone de decizie se “mută” în sus cu o cantitate α care e cu atât mai importantă cu cât raportul costurilor $\frac{C_{01}}{C_{10}}$ este mai mare. În consecință, sistemul va favoriza decizia D_0 în detrimentul deciziei D_1 (pe măsură ce dreapta “luneacă” în sus, din ce în ce mai multe puncte intră în Δ_0), ceea ce este exact comportamentul așteptat! Trebuie remarcat că, în acest caz, probabilitatea globală de eroare va crește, ceea ce scade, însă, este valoarea medie a costului.

6.3 Criteriul Bayes: cazul observațiilor continue

În acest paragraf, vom vedea ce devine regula de decizie Bayes (6.14) în cazul în care considerăm spațiul observațiilor ca fiind continuu, cu alte cuvinte, luând în considerare întregul semnal $r(t)$ pentru $t \in [0, T]$.

Vom trata problema într-un caz special, și anume considerând $s_0(t) = 0$ pentru $t \in [0, T]$. În acest caz, vom nota $s_1(t) \stackrel{\text{not}}{=} s(t)$, pentru $t \in [0, T]$, semnal având energia:

$$E = \int_0^T s^2(t) dt.$$

Observația pe baza căreia se ia decizia, respectiv semnalul $r(t)$, cu $t \in [0, T]$ poate fi văzut ca o realizare particulară a unui vector aleator dintr-un spațiu cu o infinitate nenumărabilă de dimensiuni. Vom încerca să-l reprezentăm pe acesta ca un element dintr-un spațiu cu un număr numărabil de dimensiuni. Aceasta poate fi făcută prin descompunerea lui $r(t)$ într-o bază de funcții pe $[0, T]$. Trebuie ales, deci, setul de funcții ortonormale $\{v_k(t)\}_{k \in \mathbb{N}^*}$, pentru $t \in [0, T]$, astfel încât să putem scrie:

$$r(t) = \sum_{k=1}^{\infty} r_k v_k(t), \quad t \in [0, T]. \quad (6.29)$$

Astfel, odată alese funcțiile $v_k(t)$, va exista o corespondență biunivocă între semnalul $r(t)$ și mulțimea coeficienților dezvoltării sale în baza de funcții aleasă, care este o mulțime *numărabilă*:

$$r(t), \quad t \in [0, T] \Leftrightarrow \{r_1, r_2, r_3, \dots\}.$$

Constrângerea de ortonormalitate pe care o impunem bazei $\{v_k(t)\}$, ceea ce se poate scrie compact ca:

$$\int_0^T v_k(t) v_l(t) dt = \begin{cases} 1 & \text{dacă } k = l \\ 0 & \text{dacă } k \neq l \end{cases}, \quad (6.30)$$

permite calculul coeficienților dezvoltării prin produs scalar între semnalul dezvoltat și funcția corespunzătoare din bază. Într-adevăr, putem scrie:

$$\int_0^T r(t) v_i(t) dt = \int_0^T \left(\sum_{k=1}^{\infty} r_k v_k(t) \right) v_i(t) dt = \sum_{k=1}^{\infty} r_k \underbrace{\int_0^T v_k(t) v_i(t) dt}_{=0 \text{ pentru } k \neq i} = r_i. \quad (6.31)$$

Considerentele care stau la baza alegerii setului de funcții $\{v_k(t)\}$ sunt compactarea cât mai bună a informației necesare deciziei. Cu alte cuvinte urmărim să concentrăm această informație într-un număr cât mai mic de coeficienți ai dezvoltării r_k , astfel încât decizia să se ia cât mai simplu. Se arată că, printr-o alegere potrivită a bazei, putem

face ca întreaga informație aferentă deciziei să fie concentrată într-un *singur* coeficient al dezvoltării. Astfel, alegem:

$$v_1(t) = \frac{1}{\sqrt{E}}s(t),$$

iar restul funcțiilor $v_k(t)$ cu $k = 2, 3, 4, \dots$ le alegem arbitrar, sub singura constângere de ortonormalitate (6.30)⁷.

Cu baza astfel aleasă, avem:

$$r_1 = \int_0^T r(t)v_1(t)dt = \frac{1}{\sqrt{E}} \int_0^T r(t)s(t)dt, \quad (6.32)$$

care poate fi văzut ca o realizare particulară a unei variabile aleatoare notate R_1 . În ipoteza transmiterii fiecăruia dintre cele două simboluri, avem:

$$R_1|S_0 = \int_0^T r(t)v_1(t)dt = \frac{1}{\sqrt{E}} \int_0^T n(t)s(t)dt \stackrel{\text{not}}{=} L_1, \quad (6.33)$$

$$\begin{aligned} R_1|S_1 &= \int_0^T r(t)v_1(t)dt = \frac{1}{\sqrt{E}} \int_0^T (s(t) + n(t))s(t)dt = \\ &= \underbrace{\frac{1}{\sqrt{E}} \int_0^T s^2(t)dt}_E + \underbrace{\frac{1}{\sqrt{E}} \int_0^T n(t)s(t)dt}_{L_1} = \sqrt{E} + L_1. \end{aligned} \quad (6.34)$$

Coeficientul r_k al dezvoltării pentru $k \neq 1$ poate fi văzut la rândul său ca fiind o realizare particulară a unei variabile aleatoare R_k , a cărei densitate de probabilitate poate fi scrisă, în fiecare dintre cele două ipoteze, ca:

$$R_k|S_0 = \frac{1}{\sqrt{E}} \int_0^T n(t)v_k(t)dt \stackrel{\text{not}}{=} L_k, \quad (6.35)$$

$$\begin{aligned} R_k|S_1 &= \frac{1}{\sqrt{E}} \int_0^T (s(t) + n(t))v_k(t)dt = \\ &= \underbrace{\frac{1}{\sqrt{E}} \int_0^T s(t)v_k(t)dt}_0 + \underbrace{\frac{1}{\sqrt{E}} \int_0^T n(t)v_k(t)dt}_{L_k} = L_k. \end{aligned} \quad (6.36)$$

⁷O astfel de bază există. Judecând într-un spațiu cu număr redus de dimensiuni, să zicem $N = 3$, pentru orice vector $\mathbf{s} \in \mathbb{R}^3$ se poate alege o bază ortonormală a lui $\mathbb{R}^3$, $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, în care unul din vectori să fie paralel cu $\mathbf{s}$. Astfel, se alege $\mathbf{v}_1 = \frac{\mathbf{s}}{\|\mathbf{s}\|}$, iar $\mathbf{v}_2$ și $\mathbf{v}_3$ se aleg în planul perpendicular pe $\mathbf{v}_1$, de normă unitară, și perpendiculari unul pe celălalt. Procedul de construcție a unei astfel de baze ortonormale se numește metoda Gram-Schmidt.

În dezvoltarea (6.36), s-a folosit faptul că funcțiile $v_k(t)$ cu $k \neq 1$ s-au ales ortogonale pe $v_1(t)$, deci și pe $s(t)$. Din relațiile (6.35) și (6.36), se observă cu ușurință că:

$$w_{R_k|S_0}(r_k) = w_{R_k|S_1}(r_k) \quad k = 2, 3, 4, \dots, \quad (6.37)$$

cu alte cuvinte, coeficienții $r_2, r_3, r_4, \dots$ sunt *irelevanți* pentru decizie, întrucât distribuția fiecăruia dintre ei (și, evident, a tuturor) nu depinde de simbolul transmis de sursă. Decizia se va lua numai pe baza coeficientului r_1 ⁸, iar regula de decizie Bayes (6.14) se scrie în acest caz:

$$\frac{w_{R_1|S_1}(r_1)}{w_{R_1|S_0}(r_1)} \underset{D_0}{\overset{D_1}{\gtrless}} \frac{P_0(C_{01} - C_{00})}{P_1(C_{10} - C_{11})}. \quad (6.38)$$

cu r_1 dat de (6.32).

6.3.1 Cazul zgomotului alb gaussian

Vom particulariza acum relația (6.38) în ipoteza unui zgomot alb, gaussian (același model statistic folosit în paragraful 6.2.2).

Având în vedere relațiile (6.33) și (6.34), rezultă că pentru deducerea celor două distribuții din (6.38) trebuie calculată distribuția lui

$$L_1 = \frac{1}{\sqrt{E}} \int_0^T n(t)s(t)dt. \quad (6.39)$$

Întrucât pentru $\forall t \in [0, T]$, $n(t)$ este o variabilă aleatoare gaussiană, rezultă că și $n(t)s(t)$ este o variabilă aleatoare gaussiană (rezultă din înmulțirea lui $n(t)$ cu o constantă), deci și suma tuturor acestor variabile aleatoare pentru toți $t \in [0, T]$, cu alte cuvinte însăși variabila L_1 are și ea o distribuție gaussiană: $L_1: \mathcal{N}(\overline{L_1}, \sigma_{L_1})$. Singurele nedeterminate sunt media $\overline{L_1}$ și dispersia σ_{L_1} , pe care le vom calcula în continuare. Avem:

$$\overline{L_1} = \overline{\frac{1}{\sqrt{E}} \int_0^T n(t)s(t)dt} = \frac{1}{\sqrt{E}} \int_0^T \overline{n(t)s(t)}dt = \frac{1}{\sqrt{E}} \int_0^T \underbrace{\overline{n(t)}}_0 s(t)dt = 0, \quad (6.40)$$

respectiv:

$$\begin{aligned} \overline{L_1^2} &= \overline{\left(\frac{1}{\sqrt{E}} \int_0^T n(t)s(t)dt \right)^2} = \overline{\frac{1}{E} \int_0^T \int_0^T n(t_1)s(t_1)n(t_2)s(t_2)dt_1dt_2} \\ &= \frac{1}{E} \int_0^T \int_0^T \underbrace{\overline{n(t_1)n(t_2)}}_{R_n(t_1, t_2)} s(t_1)s(t_2)dt_1dt_2 = \frac{1}{E} \int_0^T \int_0^T R_n(t_1 - t_2)s(t_1)s(t_2)dt_1dt_2. \end{aligned} \quad (6.41)$$

⁸ r_1 reprezintă statistica suficientă în acest caz.

Cum, însă, $n(t)$ este zgomot alb, conform (5.81) funcția lui de autocorelație e dată de:

$$R_n(\tau) = \sigma_n^2 \delta(\tau), \quad (6.42)$$

de unde relația (6.41) devine:

$$\begin{aligned} \overline{L_1^2} &= \frac{\sigma_n^2}{E} \int_0^T \int_0^T \delta(t_1 - t_2) s(t_1) s(t_2) dt_1 dt_2 = \frac{\sigma_n^2}{E} \int_0^T s(t_2) \underbrace{\left(\int_0^T s(t_1) \delta(t_1 - t_2) dt_1 \right)}_{s(t_2)} dt_2 \\ &= \frac{\sigma_n^2}{E} \underbrace{\int_0^T s^2(t_2) dt_2}_E = \sigma_n^2. \end{aligned} \quad (6.43)$$

În calculul de mai sus, s-a folosit relația (A.5) demonstrată în anexa A.

În continuare, din (3.21), avem:

$$\sigma_{L_1}^2 = \overline{L_1^2} - \overline{L_1}^2 = \sigma_n^2, \quad (6.44)$$

de unde rezultă, în sfârșit, că $L_1: \mathcal{N}(0, \sigma_n)$. De aici, ținând cont de (6.33) și (6.34), rezultă că $R_1|S_0: \mathcal{N}(0, \sigma_n)$, respectiv $R_1|S_1: \mathcal{N}(\sqrt{E}, \sigma_n)$, și, deci, raportul de plauzibilitate devine:

$$\begin{aligned} \Lambda(r_1) &= \frac{w_{R_1|S_1}(r_1)}{w_{R_1|S_0}(r_1)} = \frac{\frac{1}{\sigma_n \sqrt{2\pi}} \exp \left[-\frac{(r_1 - \sqrt{E})^2}{2\sigma_n^2} \right]}{\frac{1}{\sigma_n \sqrt{2\pi}} \exp \left[-\frac{r_1^2}{2\sigma_n^2} \right]} \\ &= \exp \left[-\frac{1}{2\sigma_n^2} \left((r_1 - \sqrt{E})^2 - r_1^2 \right) \right] = \exp \left[\frac{r_1 \sqrt{E}}{\sigma_n^2} - \frac{E}{2\sigma_n^2} \right], \end{aligned} \quad (6.45)$$

iar relația (6.38) se scrie, în formă logaritmică:

$$\frac{r_1 \sqrt{E}}{\sigma_n^2} - \frac{E}{2\sigma_n^2} \underset{D_0}{\overset{D_1}{\geq}} \ln K, \quad (6.46)$$

ceea ce, ținând cont de (6.32), se scrie în formă finală ca:

$$\int_0^T r(t) s(t) dt \underset{D_0}{\overset{D_1}{\geq}} \sigma_n^2 \ln K + \frac{E}{2}, \quad (6.47)$$

Din punct de vedere practic, blocul de decizie poate fi implementat folosind un filtru adaptat la semnalul $s(t)$, după cum este ilustrat în figura 6.6.

Într-adevăr, dacă se calculează valoarea semnalului la ieșirea filtrului, avem:

$$y(t) = r(t) \star h(t) = \int_{-\infty}^{\infty} r(\tau) h(t - \tau) d\tau = \int_0^t r(\tau) s(t - T + \tau) d\tau, \quad (6.48)$$

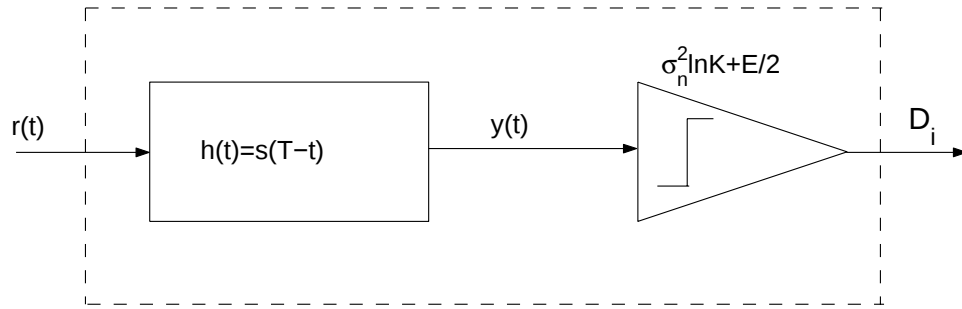


Figura 6.6: Schema bloc a blocului de detecție pentru cazul continuu.

de unde:

$$y(T) = \int_0^T r(\tau) s(\tau) d\tau.$$

Se observă că ieșirea filtrului la momentul $t = T$ reprezintă exact termenul stâng al relației (6.47). La momentul respectiv se ia decizia pe baza comparației cu termenul din dreapta (care este o constantă) după care se trece la următorul simbol etc.

Capitolul 7

Estimarea parametrilor

7.1 Introducere

Problema abordată în capitolul anterior, în cadrul detecției semnalelor, este ca pe baza măsurătorii în zgomot a unui semnal despre care știm că provine dintr-o mulțime finită de semnale cunoscute, să decidem care dintre semnalele posibile a fost emis pe canal. În cazul estimării parametrilor, dispunem de o măsurătoare în zgomot a unui semnal *cunoscut* care depinde de un parametru *necunoscut*, despre care avem doar un model statistic a priori (adică în absența oricărei măsurători). Problema estimării parametrilor, pe care o vom aborda în continuare, este ca, pe baza semnalului observat în zgomot și a modelului statistic pentru parametrul necunoscut, să-l estimăm pe acesta cât mai bine. Situația este prezentată schematic în figura 7.1.

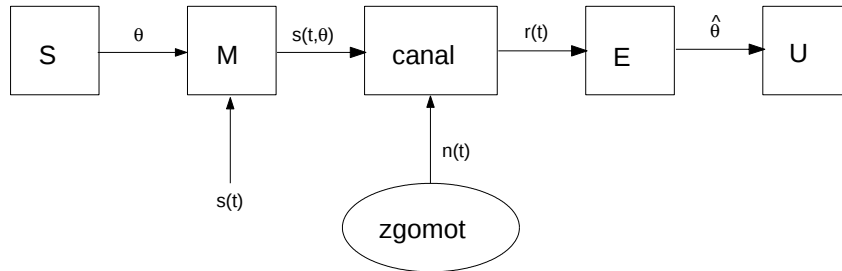


Figura 7.1: Problema estimării parametrilor.

Informația de interes este valoarea necunoscută θ , modelată ca fiind o *realizare particulară* a unei variabile aleatoare Θ , de distribuție w_Θ cunoscută:

$$\theta \equiv \Theta^{(k)}. \quad (7.1)$$

Necunoscuta θ modulează în blocul de modulare **M** în intervalul de timp $[0, T]$ un semnal cunoscut $s(t)$. Semnalul măsurat la ieșirea din canal $r(t)$ este o versiune afectată de zgomot aditiv $n(t)$ a semnalului $s(t, \theta)$, care depinde în mod cunoscut de parametrul necunoscut θ . Problema blocului de estimare **E**, pe care o vom dezvolta în continuarea

acestui capitol, este de a calcula un estimat cât mai precis al valorii necunoscute θ , estimat pe care-l notăm cu $\hat{\theta}$.

Vom discuta problema estimării numai în cazul observațiilor discrete. Cu alte cuvinte, calculul estimatului $\hat{\theta}$ se va face pe baza vectorului recepționat $\mathbf{r} = (r_1, r_2, \dots, r_N)$ format din N eșantioane ale semnalului recepționat $r(t)$, după cum s-a discutat în detaliu în paragraful 6.1. Evident, $\mathbf{r}$ este o realizare particulară a unui vector aleator N -dimensional, pe care-l vom nota, ca și în capitolul anterior, cu $\mathbf{R} = (R_1, R_2, \dots, R_N)$. De asemenea, vom păstra notațiile (6.3).

7.2 Estimarea în sensul costului mediu minim

Regula de calcul al estimatului $\hat{\theta}$ trebuie făcută în sensul minimizării unei funcții de cost C , care trebuie să depindă de eroarea de estimare ε_θ , adică de diferența dintre estimat și valoarea reală:

$$C = C(\varepsilon_\theta)$$

unde

$$\varepsilon_\theta = \theta - \hat{\theta}.$$

Evident, pentru ca funcția C să îndeplinească semnificația de cost al estimării, ea trebuie aleasă crescătoare cu modulul argumentului. Cu alte cuvinte, costul estimării trebuie să crească atunci când eroarea de estimare crește în valoare absolută.

Întrucât valoarea estimată $\hat{\theta}$ este o funcție de cele N valori ale semnalului observat, rezultă că funcția de cost C este o funcție de $N+1$ variabile aleatoare, respectiv parametrul ce trebuie estimat și cele N eșantioane ale semnalului recepționat. Rezultă că valoarea medie a costului $\overline{C}$ se va calcula prin extensia directă a teoremei de medie (4.30) la cazul $N+1$ dimensional:

$$\overline{C} = \int_{\mathbb{R}^{N+1}} C(\theta - \hat{\theta}) w_{\mathbf{R}, \Theta}(\mathbf{r}, \theta) d\mathbf{r} d\theta. \quad (7.2)$$

În continuare, vom căuta să calculăm estimatul $\hat{\theta}$ astfel încât costul mediu al estimării (7.2) să fie minim.

Astfel, scriem densitatea de probabilitate de ordinul $N+1$ conform (4.16):

$$w_{\mathbf{R}, \Theta}(\mathbf{r}, \theta) = w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}(\theta) w_{\mathbf{R}}(\mathbf{r}) = w_{\mathbf{R}|\{\Theta=\theta\}}(\mathbf{r}) w_{\Theta}(\theta) \quad (7.3)$$

după care costul mediu poate fi scris, separând integrala $N+1$ dimensională într-o integrală unidimensională după θ , urmată de una N dimensională după $\mathbf{r}$:

$$\overline{C} = \int_{\mathbb{R}^N} w_{\mathbf{R}}(\mathbf{r}) \left(\int_{-\infty}^{\infty} C(\theta - \hat{\theta}) w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}(\theta) d\theta \right) d\mathbf{r} \quad (7.4)$$

Vom nota integrala după θ cu I , care, evident, este o funcție atât de vectorul $\mathbf{r}$, cât și de $\hat{\theta}$:

$$I(\mathbf{r}, \hat{\theta}) \stackrel{\text{not}}{=} \int_{-\infty}^{\infty} C(\theta - \hat{\theta}) w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}(\theta) d\theta. \quad (7.5)$$

Costul mediu $\bar{C}$ devine minim dacă pentru fiecare observație $\mathbf{r}$ alegem $\hat{\theta}$ astfel încât integrala $I(\mathbf{r}, \hat{\theta})$ din (7.5) să fie minimă.

În paragrafele următoare, vom deduce formula finală a estimatului care minimizează pe I pentru două funcții de cost precizate.

7.2.1 Estimarea în sensul funcției de cost pătratul erorii

Considerăm în acest paragraf cazul funcției de cost pătratice:

$$C_p(\varepsilon_\theta) = \varepsilon_\theta^2. \quad (7.6)$$

Alegerea funcției pătratice pentru definirea costului estimării implică faptul că vom “insera” în regula de calcul a estimatului o informație conform căreia cu cât eroarea de estimate este mai mare, cu atâta costul estimării este mai important!

Înlocuind (7.6) în (7.5), obținem:

$$I(\mathbf{r}, \hat{\theta}) = \int_{-\infty}^{\infty} (\theta - \hat{\theta})^2 w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}(\theta) d\theta. \quad (7.7)$$

Valoarea lui $\hat{\theta}$ care minimizează pe I se obține din ecuația:

$$\frac{\partial I(\mathbf{r}, \hat{\theta})}{\partial \hat{\theta}} = 0. \quad (7.8)$$

Astfel, avem:

$$\begin{aligned} \frac{\partial I(\mathbf{r}, \hat{\theta})}{\partial \hat{\theta}} &= -2 \int_{-\infty}^{\infty} (\theta - \hat{\theta}) w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}(\theta) d\theta \\ &= 2\hat{\theta} \underbrace{\int_{-\infty}^{\infty} w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}(\theta) d\theta}_1 - 2 \int_{-\infty}^{\infty} \theta w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}(\theta) d\theta. \end{aligned} \quad (7.9)$$

În dezvoltarea de mai sus, s-a ținut cont de cele precizate în paragraful 3.5, conform cărora o densitate de probabilitate condiționată respectă toate proprietățile unei densități de probabilitate, deci și condiția de normare. Astfel, aplicând (7.8) obținem expresia finală a estimatului optim în sensul funcției de cost pătratul erorii, numit *estimat pătratic* și notat $\hat{\theta}_p$:

$$\hat{\theta}_p = \int_{-\infty}^{\infty} \theta w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}(\theta) d\theta \quad (7.10)$$

Interpretarea relației (7.10) este că estimatul pătratic al unui parametru se calculează ca media condiționată a posteriori a acestuia, respectiv media statistică a parametrului în condițiile în care vectorul observat este chiar $\mathbf{r}$.

7.2.2 Estimarea în sensul funcției de cost uniforme

În acest paragraf, vom deduce formula estimatului optim în sensul funcției de cost uniforme, definită ca:

$$C_u(\varepsilon_\theta) = \begin{cases} 0 & \text{dacă } |\varepsilon_\theta| \leq \frac{E}{2} \\ 1 & \text{în rest} \end{cases}. \quad (7.11)$$

Alegerea funcției (7.11) pe post de cost al estimării este echivalentă cu a spune că nu ne interesează decât ca eroarea de estimare să fie între niște limite bine precizate. Atâta vreme cât eroarea este mai mică decât limita respectivă, costul este nul, în timp ce dacă limita a fost depășită, nu ne interesează cu cât anume a fost depășită.

Rescriind relația (7.11) ca:

$$C_u(\varepsilon_\theta) = 1 - \begin{cases} 1 & \text{dacă } \hat{\theta} - \frac{E}{2} \leq \theta \leq \hat{\theta} + \frac{E}{2} \\ 0 & \text{în rest} \end{cases}, \quad (7.12)$$

și înlocuind-o în (7.5), obținem:

$$I(\mathbf{r}, \hat{\theta}) = \underbrace{\int_{-\infty}^{\infty} w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}(\theta) d\theta}_1 - \int_{\hat{\theta}-\frac{E}{2}}^{\hat{\theta}+\frac{E}{2}} w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}(\theta) d\theta, \quad (7.13)$$

de unde rezultă că minimizarea lui I este echivalentă cu maximizarea după $\hat{\theta}$ a integralei:

$$I'(\mathbf{r}, \hat{\theta}) = \int_{\hat{\theta}-\frac{E}{2}}^{\hat{\theta}+\frac{E}{2}} w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}(\theta) d\theta. \quad (7.14)$$

Problema maximizării lui I' din (7.14) nu poate fi rezolvată analitic decât printr-o aproximare suplimentară, și anume făcând ipoteza că lățimea E a intervalului de acceptabilitate a erorii de estimare este foarte mică. Atunci, putem scrie:

$$w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}(\theta) \underset{E \searrow 0}{\approx} \text{constant} = w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}(\hat{\theta}), \quad \forall \theta \in \left[\hat{\theta} - \frac{E}{2}, \hat{\theta} + \frac{E}{2} \right] \quad (7.15)$$

ceea ce conduce la:

$$I'(\mathbf{r}, \hat{\theta}) \approx w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}(\hat{\theta}) \underbrace{\int_{\hat{\theta}-\frac{E}{2}}^{\hat{\theta}+\frac{E}{2}} d\theta}_E = E w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}(\hat{\theta}). \quad (7.16)$$

Astfel, maximizarea lui I' revine la maximizarea densității de probabilitate $w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}(\theta)$. Se numește estimat *maximum a posteriori* și se notează cu $\hat{\theta}_{\text{map}}$ estimatul optim în sensul funcției de cost uniforme, și este acea valoare a lui θ care maximizează densitatea de probabilitate a posteriori:

$$w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}(\theta) \Big|_{\theta=\hat{\theta}_{\text{map}}} = \max. \quad (7.17)$$

În figura 7.2 este ilustrată grafic diferența între cei doi estimatori.

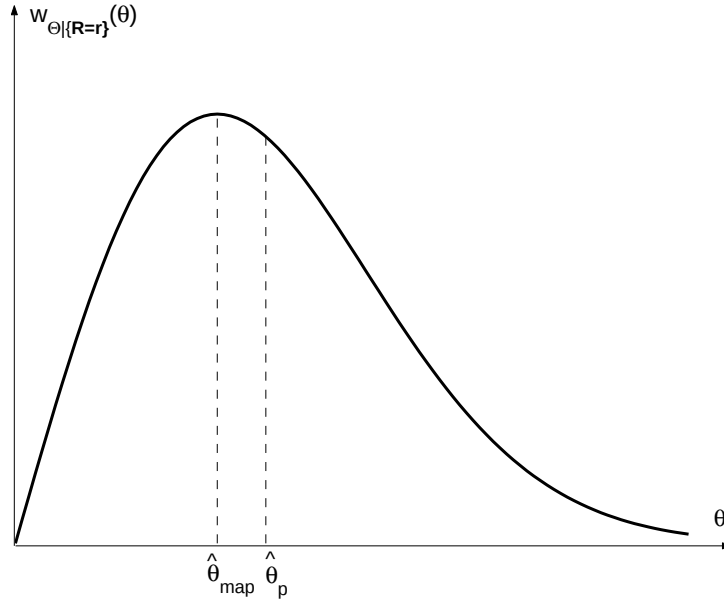


Figura 7.2: Diferența între estimatul pătratic și cel maximum a posteriori pentru o distribuție a posteriori $w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}$ oarecare.

Relația (7.17) poate fi pusă într-o formă în care intervin densitatea de probabilitate a priori $w_{\mathbf{R}|\{\Theta=\theta\}}(\mathbf{r})$ al cărei calcul este mai simplu de făcut. În primul rând, exprimăm relația (7.17) în formă logaritmică:

$$\ln w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}(\theta) \Big|_{\theta=\hat{\theta}_{\text{map}}} = \max, \quad (7.18)$$

ceea ce este echivalent cu a spune că:

$$\frac{\partial}{\partial \theta} \ln w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}(\theta) \Big|_{\theta=\hat{\theta}_{\text{map}}} = 0 \quad (7.19)$$

Din (7.3), scriem densitatea a posteriori a lui Θ ca:

$$w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}(\theta) = \frac{w_{\mathbf{R}|\{\Theta=\theta\}}(\mathbf{r})w_{\Theta}(\theta)}{w_{\mathbf{R}}(\mathbf{r})}, \quad (7.20)$$

de unde (7.19) se scrie:

$$\left. \frac{\partial}{\partial \theta} \ln \left(\frac{w_{\mathbf{R}|\{\Theta=\theta\}}(\mathbf{r}) w_{\Theta}(\theta)}{w_{\mathbf{R}}(\mathbf{r})} \right) \right|_{\theta=\hat{\theta}_{\text{map}}} = 0. \quad (7.21)$$

Ținând cont de proprietatea logaritmului de a transforma produsul în sumă, putem scrie relația de mai sus sub forma:

$$\left(\frac{\partial}{\partial \theta} \ln w_{\mathbf{R}|\{\Theta=\theta\}}(\mathbf{r}) + \frac{\partial}{\partial \theta} \ln w_{\Theta}(\theta) - \frac{\partial}{\partial \theta} \ln w_{\mathbf{R}}(\mathbf{r}) \right) \Big|_{\theta=\hat{\theta}_{\text{map}}} = 0, \quad (7.22)$$

de unde, observând că ultimul termen din partea stângă a ecuației este nul (densitatea $w_{\mathbf{R}}(\mathbf{r})$ nu depinde de θ) obținem forma finală a ecuației ce permite determinarea lui $\hat{\theta}_{\text{map}}$:

$$\left(\frac{\partial}{\partial \theta} \ln w_{\mathbf{R}|\{\Theta=\theta\}}(\mathbf{r}) + \frac{\partial}{\partial \theta} \ln w_{\Theta}(\theta) \right) \Big|_{\theta=\hat{\theta}_{\text{map}}} = 0. \quad (7.23)$$

În paragraful următor, vom prezenta un exemplu de calcul al estimatului unui parametru pentru ipoteze statistice bine precizate pentru acesta și pentru zgomot.

7.2.3 Estimarea unui parametru gaussian în zgomot alb, gaussian

Vom presupune că θ provine dintr-o distribuție gaussiană, de medie $\bar{\Theta}$ și dispersie σ_{Θ} cunoscute: $\Theta: \mathcal{N}(\bar{\Theta}, \sigma_{\Theta})$. De asemenea, vom considera pentru zgomotul $n(t)$ același model statistic folosit în paragraful 6.2.2, și anume zgomot alb, cu distribuție normală de medie 0 și dispersie σ_n .

Vom presupune în plus că semnalul modulat depinde în mod liniar de θ :

$$s(t, \theta) = \theta s(t), \quad t \in [0, T], \quad (7.24)$$

cu $s(t)$ un semnal cunoscut. În aceste condiții, semnalul recepționat este:

$$r(t) = s(t, \theta) + n(t) = \theta s(t) + n(t), \quad t \in [0, T]. \quad (7.25)$$

Să precizăm în continuare că, deși problema calculului estimatului lui θ poate fi rezolvată direct, și mai simplu, aplicând una din formulele (7.10) sau (7.23), noi vom proceda prin calculul intermediar al distribuției a posteriori $w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}(\theta)$ a parametrului necunoscut, considerând că astfel se poate ajunge la o mai bună înțelegere a problemei.

Pornim calculul densității de probabilitate a posteriori de la relația (7.20). În ceea ce privește distribuția globală a vectorului recepționat $w_{\mathbf{R}}(\mathbf{r})$, o putem scrie ca o densitate marginală a lui $w_{\mathbf{R},\Theta}(\mathbf{r}, \theta)$ aplicând extensia relației (4.6) pentru cazul multidimensional:

$$w_{\mathbf{R}}(\mathbf{r}) = \int_{-\infty}^{\infty} w_{\mathbf{R},\Theta}(\mathbf{r}, \theta) d\theta = \int_{-\infty}^{\infty} w_{\mathbf{R}|\{\Theta=\theta\}}(\mathbf{r}) w_{\Theta}(\theta) d\theta. \quad (7.26)$$

Rezultă, deci, că:

$$w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}(\theta) = \frac{w_{\mathbf{R}|\{\Theta=\theta\}}(\mathbf{r})w_{\Theta}(\theta)}{\int_{-\infty}^{\infty} w_{\mathbf{R}|\{\Theta=\theta\}}(\mathbf{r})w_{\Theta}(\theta)d\theta}, \quad (7.27)$$

relație ce face legătura între densitatea de probabilitate a posteriori $w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}(\theta)$ și densitățile a priori $w_{\mathbf{R}|\{\Theta=\theta\}}(\mathbf{r})$, respectiv $w_{\Theta}(\theta)$, densități ce pot fi calculate direct pe baza ipotezelor statistice asumate.

Pentru calculul densității de probabilitate a vectorului $\mathbf{R}$ în ipoteza $\Theta = \theta$, pornim de la (7.25), de unde rezultă că eșantionul $r(t_i)$ al semnalului recepționat este realizarea particulară a variabilei R_i , care, în ipoteza transmiterii lui θ se scrie:

$$R_i|\{\Theta = \theta\} = \theta s_i + n_i, \quad i = 1, \dots, N, \quad (7.28)$$

unde $s_i \stackrel{\text{not}}{=} s(t_i)$ și $n_i \stackrel{\text{not}}{=} n(t_i)$.

Astfel, din (7.28), rezultă că $R_i|\{\Theta = \theta\} : \mathcal{N}(\theta s_i, \sigma_n)$, de unde, pe baza independenței între eșantioanele semnalului $r(t)$ (conform celor discutate în detaliu în paragraful 6.2.2), avem:

$$w_{\mathbf{R}|\{\Theta=\theta\}}(\mathbf{r}) = \prod_{i=1}^N w_{R_i|\{\Theta=\theta\}}(r_i) = \left(\frac{1}{\sigma_n \sqrt{2\pi}} \right)^N \exp \left[-\frac{1}{2\sigma_n^2} \sum_{i=1}^N (r_i - \theta s_i)^2 \right]. \quad (7.29)$$

Cum, tot conform ipotezelor inițiale,

$$w_{\Theta}(\theta) = \frac{1}{\sigma_{\Theta} \sqrt{2\pi}} \exp \left[-\frac{(\theta - \bar{\Theta})^2}{2\sigma_{\Theta}^2} \right], \quad (7.30)$$

rezultă că (7.27) se scrie:

$$w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}(\theta) = \frac{\left(\frac{1}{\sigma_n \sqrt{2\pi}} \right)^N \frac{1}{\sigma_{\Theta} \sqrt{2\pi}} \exp \left[-\frac{1}{2\sigma_n^2} \sum_{i=1}^N (r_i - \theta s_i)^2 \right] \exp \left[-\frac{(\theta - \bar{\Theta})^2}{2\sigma_{\Theta}^2} \right]}{\left(\frac{1}{\sigma_n \sqrt{2\pi}} \right)^N \frac{1}{\sigma_{\Theta} \sqrt{2\pi}} \int_{-\infty}^{\infty} \exp \left[-\frac{1}{2\sigma_n^2} \sum_{i=1}^N (r_i - \theta s_i)^2 \right] \exp \left[-\frac{(\theta - \bar{\Theta})^2}{2\sigma_{\Theta}^2} \right] d\theta}, \quad (7.31)$$

ceea ce, după reduceri și gruparea termenilor, poate fi scris sub formă compactă ca:

$$w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}(\theta) = \frac{\exp(-A\theta^2 + B\theta - C)}{\int_{-\infty}^{\infty} \exp(-A\theta^2 + B\theta - C)d\theta} \quad (7.32)$$

cu

$$A \stackrel{\text{not}}{=} \frac{\sum_{i=1}^N s_i^2}{2\sigma_n^2} + \frac{1}{2\sigma_{\Theta}^2} \quad (7.33a)$$

$$B \stackrel{\text{not}}{=} \frac{\sum_{i=1}^N r_i s_i}{\sigma_n^2} + \frac{\bar{\Theta}}{\sigma_{\Theta}^2} \quad (7.33b)$$

$$C \stackrel{\text{not}}{=} \frac{\sum_{i=1}^N r_i^2}{2\sigma_n^2} + \frac{\bar{\Theta}^2}{2\sigma_{\Theta}^2} \quad (7.33c)$$

Punând forma pătratică ce constituie argumentul exponențialei din (7.32) sub forma ei canonică, avem:

$$\exp(-A\theta^2 + B\theta - C) = \exp\left[-A\left(\theta - \frac{B}{2A}\right)^2\right] \underbrace{\exp\left(\frac{B^2}{4A} - C\right)}_{\text{const}}. \quad (7.34)$$

Astfel, relația (7.32) devine:

$$\begin{aligned} w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}(\theta) &= \frac{\exp\left(\frac{B^2}{4A} - C\right) \exp\left[-A\left(\theta - \frac{B}{2A}\right)^2\right]}{\exp\left(\frac{B^2}{4A} - C\right) \int_{-\infty}^{\infty} \exp\left[-A\left(\theta - \frac{B}{2A}\right)^2\right] d\theta} \\ &= \frac{1}{\sqrt{\frac{\pi}{A}}} \exp\left[-A\left(\theta - \frac{B}{2A}\right)^2\right]. \end{aligned} \quad (7.35)$$

În relația (7.35) s-a ținut cont că pentru $\forall \alpha > 0$ și $\forall t_0 \in \mathbb{R}$ avem:

$$\int_{-\infty}^{\infty} \exp[-\alpha(t - t_0)^2] dt = \sqrt{\frac{\pi}{\alpha}}. \quad (7.36)$$

Confruntând (7.35) cu (3.23), putem scrie:

$$\Theta|\{\mathbf{R} = \mathbf{r}\}: \mathcal{N}\left(\frac{B}{2A}, \frac{1}{\sqrt{2A}}\right). \quad (7.37)$$

Cu alte cuvinte, distribuția lui Θ *a posteriori*, ținând cont de informația observată și conținută în $\mathbf{r}$, este tot gaussiană, având media și dispersia date de:

$$\bar{\Theta}|\{\mathbf{R} = \mathbf{r}\} = \frac{B}{2A} = \frac{\frac{\sum_{i=1}^N r_i s_i}{\sigma_n^2} + \frac{\bar{\Theta}}{\sigma_{\Theta}^2}}{\frac{\sum_{i=1}^N s_i^2}{\sigma_n^2} + \frac{1}{\sigma_{\Theta}^2}}, \quad (7.38)$$

$$\sigma_{\Theta|\{\mathbf{R}=\mathbf{r}\}} = \frac{1}{\sqrt{2A}} = \frac{1}{\sqrt{\frac{\sum_{i=1}^N s_i^2}{\sigma_n^2} + \frac{1}{\sigma_{\Theta}^2}}}. \quad (7.39)$$

Să interpretăm rezultatul obținut într-un caz particular, și anume $s(t) = 1$, pentru $t \in [0, T]$. În acest caz, relația (7.28) devine:

$$R_i|\{\Theta = \theta\} = \theta + n_i, \quad i = 1, \dots, N. \quad (7.40)$$

Altfel spus, dispunem de N observații ale lui θ afectate de zgomot, din care trebuie să-l estimăm pe acesta. În această situație particulară, media și dispersia lui $\Theta|\{\mathbf{R} = \mathbf{r}\}$ devin:

$$\bar{\Theta}|\{\mathbf{R} = \mathbf{r}\} = \frac{\frac{\sum_{i=1}^N r_i}{\sigma_n^2} + \frac{\bar{\Theta}}{\sigma_{\Theta}^2}}{\frac{N}{\sigma_n^2} + \frac{1}{\sigma_{\Theta}^2}} = \frac{\frac{1}{N} \sum_{i=1}^N r_i}{1 + \frac{1}{N} \frac{\sigma_n^2}{\sigma_{\Theta}^2}} + \frac{\bar{\Theta}}{1 + N \frac{\sigma_{\Theta}^2}{\sigma_n^2}} \quad (7.41)$$

$$\sigma_{\Theta|\{\mathbf{R}=\mathbf{r}\}} = \frac{1}{\sqrt{\frac{N}{\sigma_n^2} + \frac{1}{\sigma_{\Theta}^2}}}. \quad (7.42)$$

În figura 7.3 sunt prezentate densitatea de probabilitate a priori $w_{\Theta}(\theta)$ și densitatea de probabilitate a posteriori $w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}(\theta)$ în două cazuri diferite, pentru valori numerice precizate ale parametrilor statistici și ale vectorului observat $\mathbf{r}$. Se observă că, față de distribuția a priori $w_{\Theta}(\theta)$, distribuția a posteriori $w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}(\theta)$ se “deplasează” în direcția dată de observația $\mathbf{r}$. Mai mult, se observă că dispersia parametrului $\Theta|\{\mathbf{R} = \mathbf{r}\}$ este întotdeauna mai mică decât dispersia lui Θ (ceea ce poate fi și demonstrat matematic din relația (7.42)) ceea ce este normal, întrucât măsurătoarea $\mathbf{r}$ de care dispunem aduce informație suplimentară asupra lui Θ .

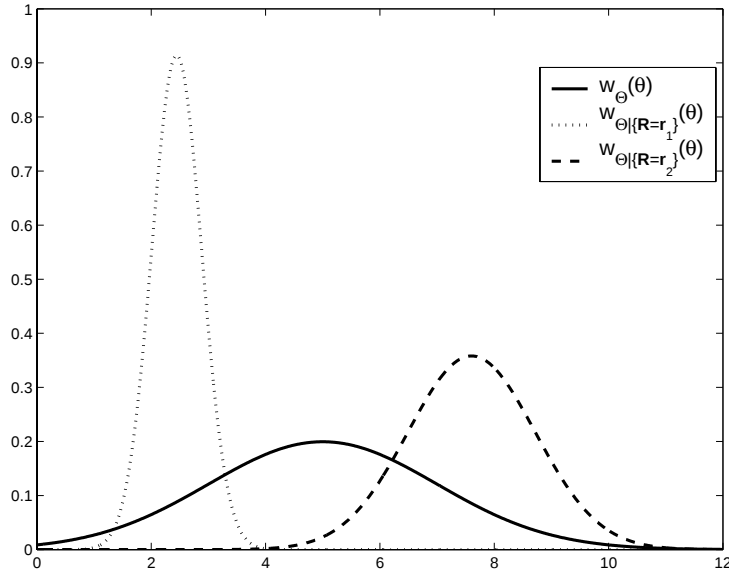


Figura 7.3: Distribuția a priori $w_{\Theta}(\theta)$ pentru $\bar{\Theta} = 5$ și $\sigma_{\Theta} = 2$ și distribuția a posteriori $w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}(\theta)$ pentru doi vectori recepționați $\mathbf{r}_1 = [1.6; 2.69; 2.82; 2.71; 3.29]$, respectiv $\mathbf{r}_2 = [9.0; 11.57; 5.39; 7.94; 7.53]$.

Evident, întrucât $w_{\Theta|\{\mathbf{R}=\mathbf{r}\}}(\theta)$ este un clopot al lui Gauss, valoarea care îl maximizează pe acesta coincide cu valoarea medie, deci estimatul pătratic este egal cu estimatul maximum a posteriori (vezi figura 7.2):

$$\hat{\theta}_p = \hat{\theta}_{\text{map}} = \bar{\Theta}|\{\mathbf{R} = \mathbf{r}\} = \frac{\frac{1}{N} \sum_{i=1}^N r_i}{1 + \frac{1}{N} \frac{\sigma_n^2}{\sigma_{\Theta}^2}} + \frac{\bar{\Theta}}{1 + N \frac{\sigma_{\Theta}^2}{\sigma_n^2}}. \quad (7.43)$$

Relația de mai sus poate fi interpretată mai ușor făcând următoarele notații:

$$\alpha \stackrel{\text{not}}{=} \frac{1}{1 + \frac{1}{N} \frac{\sigma_n^2}{\sigma_{\Theta}^2}}, \quad (7.44a)$$

$$\beta \stackrel{\text{not}}{=} \frac{1}{1 + N \frac{\sigma_{\Theta}^2}{\sigma_n^2}}. \quad (7.44b)$$

Se observă apoi că

$$\alpha + \beta = 1. \quad (7.45)$$

Notând, de asemenea, media celor N eşantioane ale lui $\mathbf{r}$ cu $\mu_N(\mathbf{r})$:

$$\mu_N(\mathbf{r}) \stackrel{\text{not}}{=} \frac{1}{N} \sum_{i=1}^N r_i, \quad (7.46)$$

şi ținând cont de (7.45), relația (7.43) se scrie:

$$\hat{\theta}_p = \hat{\theta}_{\text{map}} \stackrel{\text{not}}{=} \hat{\theta} = \alpha \mu_N(\mathbf{r}) + (1 - \alpha) \bar{\Theta}. \quad (7.47)$$

Rezultă, deci, că estimatul lui θ se obține ca o combinație convexă între doi termeni: primul este media aritmetică a observațiilor lui θ în zgomot $\mu_N(\mathbf{r})$, care reprezintă estimatul “practic” al lui θ , bazat numai pe observații, în timp ce al doilea este $\bar{\Theta}$, care reprezintă estimatul “a priori” al lui θ , bazat numai pe cunoștințele statistice a priori asupra parametrului necunoscut. Situația este ilustrată în figura 7.4.



Figura 7.4: Estimatul $\hat{\theta}$ este o combinație convexă între estimatul “practic” $\mu_N(\mathbf{r})$ și cel “a priori” $\bar{\Theta}$.

Ponderile celor doi termeni în valoarea finală a lui $\hat{\theta}$ depind de trei parametri, și anume dispersia zgomotului σ_n , dispersia cunoscută a priori a parametrului σ_Θ și numărul de observații N . Să analizăm pe scurt dependența ponderilor de fiecare din cei trei parametri (luat în considerare independent de ceilalți):

- $N \nearrow \Rightarrow \alpha \nearrow 1 \Rightarrow \hat{\theta} \rightarrow \mu_N(\mathbf{r})$. Cu alte cuvinte, pe măsură ce avem din ce în ce mai multe măsurători, tindem să dăm o pondere din ce în ce mai mare măsurătorii în detrimentul cunoștințelor a priori.
- $\sigma_n \nearrow \Rightarrow \alpha \searrow 0 \Rightarrow \hat{\theta} \rightarrow \bar{\Theta}$. Altfel spus, ponderea estimatului a priori crește cu dispersia zgomotului, care este o măsură a puterii acestuia. Acest lucru e normal întrucât cu cât zgomotul de măsurătoare este mai mare, cu atât observația este mai puțin “fiabilă”.
- $\sigma_\Theta \nearrow \Rightarrow \alpha \nearrow 1 \Rightarrow \hat{\theta} \rightarrow \mu_N(\mathbf{r})$. σ_Θ este dispersia cunoscută a priori a variabilei aleatoare Θ din care provine θ . Cu alte cuvinte σ_Θ este o măsură a cât de verosimilă este valoarea $\bar{\Theta}$ pentru θ . Este normal, deci, ca ponderea estimatului a priori să scadă cu σ_Θ . În cele două cazuri limită, putem scrie: $\hat{\theta} \xrightarrow[\sigma_\Theta \rightarrow 0]{} \bar{\Theta}$, cu alte cuvinte, măsurătoarea nu contează, întrucât știm dinainte că $\theta = \bar{\Theta}$, respectiv $\hat{\theta} \xrightarrow[\sigma_\Theta \rightarrow \infty]{} \mu_N(\mathbf{r})$. Acest din urmă caz, respectiv $\sigma_\Theta \rightarrow \infty$ modelează absența cunoștințelor a priori asupra lui θ . Într-adevăr, pe măsură ce σ_Θ crește, clopotul w_Θ se aplatizează, apropiindu-se din ce în ce mai mult de o funcție constantă pe $\mathbb{R}$. Evident, în acest caz, estimarea se face numai pe baza observației. În paragraful următor, vom da o soluție generală pentru acest caz.

7.2.4 Estimarea în absența unui model statistic a priori

După cum am prezentat în exemplul precedent, faptul că nu avem nici o cunoștință a priori asupra modelului statistic al parametrului necunoscut θ este echivalent cu a spune, în absența oricărei măsurători, că el poate lua orice valoare cu egală probabilitate, altfel spus:

$$w_{\Theta}(\theta) = \text{constant } \forall \theta \in \mathbb{R} \quad (7.48)$$

inserând (7.48) în (7.23), obținem relația de calcul al estimatului. Se numește estimat de maximă plauzibilitate și se notează cu $\hat{\theta}_{\text{mp}}$ acea valoare a lui θ care maximizează densitatea de probabilitate a priori $w_{\mathbf{R}|\{\Theta=\theta\}}(\mathbf{r})$, sau, altfel spus:

$$\left. \frac{\partial}{\partial \theta} \ln w_{\mathbf{R}|\{\Theta=\theta\}}(\mathbf{r}) \right|_{\theta=\hat{\theta}_{\text{mp}}} = 0. \quad (7.49)$$

Este de remarcat faptul că estimatul de maximă plauzibilitate nu este o alternativă la cel pătratic sau maximum a posteriori, ci este soluția într-un caz particular, și anume lipsa informațiilor statistice a priori asupra parametrului necunoscut. Problema calculului lui $\hat{\theta}_{\text{mp}}$ nu se pune dacă știm $w_{\Theta}(\theta)$!

7.3 Evaluarea calității unui estimator

Vom prezenta în continuare câteva criterii de evaluare a calității unui estimator. Evident, estimatul $\hat{\theta}$ al valorii θ date este o variabilă aleatoare. Cu alte cuvinte, repetarea estimării *aceluiași* θ va conduce de fiecare dată la valori diferite ale lui $\hat{\theta}$, indiferent de metoda de calcul aleasă. Valoarea medie a estimatului lui θ după toate realizările posibile ale vectorului aleator recepționat $\mathbf{R}|\{\Theta = \theta\}$ este dată de:

$$\bar{\hat{\theta}} = \int_{\mathbb{R}^N} \hat{\theta}(\mathbf{r}) w_{\mathbf{R}|\{\Theta=\theta\}}(\mathbf{r}) d\mathbf{r}. \quad (7.50)$$

Estimatul se numește *nedeplasat* dacă $\bar{\hat{\theta}} = \theta$. În caz contrar, estimatul se numește *deplasat*. Evident, cu cât diferența $\bar{\hat{\theta}} - \theta$ este mai mare, cu atât calitatea estimatorului scade.

De asemenea, se poate calcula și varianța estimatului $\sigma_{\hat{\theta}}$ după:

$$\sigma_{\hat{\theta}}^2 = \overline{(\hat{\theta} - \bar{\hat{\theta}})^2} = \int_{\mathbb{R}^N} (\hat{\theta}(\mathbf{r}) - \bar{\hat{\theta}})^2 w_{\mathbf{R}|\{\Theta=\theta\}}(\mathbf{r}) d\mathbf{r}. \quad (7.51)$$

Varianța (dispersia) estimatului $\hat{\theta}$ este o măsură a fluctuației valorilor acestuia în jurul valorii medii. Evident, cu cât dispersia este mai mică, cu atât calitatea estimatorului crește. În general, este de preferat un estimator deplasat cu dispersie mică unuia nedeplasat dar având dispersie mare.

Capitolul 8

Semnale aleatoare în timp discret

8.1 Introducere

În acest capitol, vom începe studiul semnalelor aleatoare în timp discret. Acestea sunt obținute din semnalele aleatoare în timp continuu prin eșantionare, cu o anumită perioadă T_e , după cum este ilustrat în figura 8.1. Conform teoremei eșantionării, pentru ca semnalul continuu să poată fi reconstituit fără nici o pierdere din eșantioanele sale, trebuie ca frecvența de eșantionare să fie de cel puțin două ori mai mare ca frecvența maximă a semnalului $f_{\max}$:

$$f_e = \frac{1}{T_e} \geq 2f_{\max}. \quad (8.1)$$

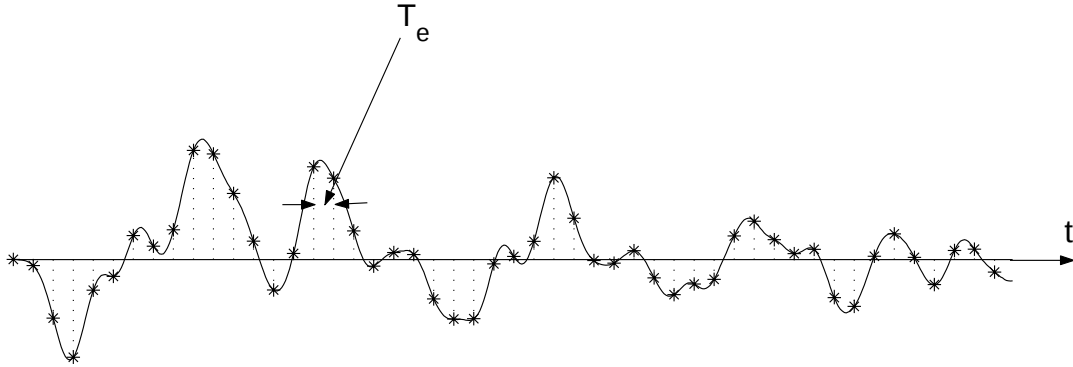


Figura 8.1: Eșantionarea semnalelor

Astfel, semnalele aleatoare în timp discret sunt reprezentate sub formă de secvență de numere $[\dots, \xi(-T_e), \xi(0), \xi(T_e), \xi(2T_e), \dots, \xi(nT_e), \dots]$ fiecare dintre aceste numere fiind o variabilă aleatoare *continuă*. În cele ce urmează, vom omite din notație perioada de eșantionare, adică vom nota:

$$\xi(n) \stackrel{\text{not}}{=} \xi(nT_e). \quad (8.2)$$

Reprezentarea semnalelor în timp discret este unul din cei doi pași care fac posibilă prelucrarea semnalelor cu calculatorul. Al doilea pas este cuantizarea (subiect pe care-l vom aborda în detaliu în capitolul 11) care implică transformarea naturii valorilor eșantioanelor $\xi(n)$ din continuă în discretă.

Evident, datorită naturii discrete a reprezentării temporale a semnalului, atât densitățile de probabilitate de orice ordin cât și momentele semnalului vor fi și ele reprezentate tot în timp discret. În rest, toate considerentele discutate în capitolul 5 rămân valabile și în reprezentarea în timp discret. În particular, un semnal aleator în timp discret $\xi(n)$ se numește staționar în sens larg (pe scurt, staționar) dacă:

$$\overline{\xi(n)} = \bar{\xi} \quad (8.3)$$

$$R_{\xi}(n, k) = \overline{\xi(n)\xi(k)} = R_{\xi}(n - k). \quad (8.4)$$

În continuare, vom presupune ipoteza de staționaritate pentru toate semnalele în timp discret considerate.

În ceea ce privește reprezentarea spectrală a semnalelor aleatoare în timp discret, sigura modificare față de reprezentarea în timp continuu este dată de faptul că, datorită eșantionării cu perioadă T_e , cea mai mare frecvență ce poate fi reprezentată în spectrul semnalului (conform condiției Nyquist) este:

$$\omega_{\max} = \frac{\omega_e}{2} = \frac{\pi}{T_e}. \quad (8.5)$$

Deci, ținând cont de convenția de notație $T_e = 1$, densitatea spectrală de putere a semnalului, obținută ca fiind transformata Fourier în timp discret a funcției sale de autocorelație:

$$q_{\xi}(\omega) = \mathcal{F}\{R_{\xi}(k)\}(\omega) = \sum_{k=-\infty}^{\infty} R_{\xi}(k) \exp(-j\omega k), \quad (8.6)$$

va avea componente în intervalul de frecvențe $[-\pi, \pi]$.

8.2 Matricea de autocorelație

În acest paragraf, vom defini o structură foarte importantă în studiul semnalelor aleatoare în timp discret, și anume matricea de autocorelație. Vom porni prin a defini vectorul $\boldsymbol{\xi} \in \mathbb{R}_{N \times 1}$ ca fiind vectorul coloană ce conține N eșantioane ale unui semnal staționar:

$$\boldsymbol{\xi} = [\xi(0) \ \xi(1) \ \cdots \ \xi(N-1)]^T. \quad (8.7)$$

unde prin $()^T$ am notat operatorul de transpunere de matrici. Matricea de autocorelație a semnalului, notată cu $\mathbf{R}_{\xi}$, se definește ca:

$$\mathbf{R}_{\xi} \triangleq \boldsymbol{\xi} \boldsymbol{\xi}^T \quad (8.8)$$

Explicitând înmulțirea și ținând cont de faptul că funcția de autocorelație a semnalului este pară, avem:

$$\begin{aligned}
 \mathbf{R}_\xi &= \begin{bmatrix} \xi(0) \\ \xi(1) \\ \vdots \\ \xi(N-1) \end{bmatrix} \begin{bmatrix} \xi(0) & \xi(1) & \cdots & \xi(N-1) \end{bmatrix} \\
 &= \begin{bmatrix} \overline{\xi(0)\xi(0)} & \overline{\xi(0)\xi(1)} & \cdots & \overline{\xi(0)\xi(N-1)} \\ \overline{\xi(1)\xi(0)} & \overline{\xi(1)\xi(1)} & \cdots & \overline{\xi(1)\xi(N-1)} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{\xi(N-1)\xi(0)} & \overline{\xi(N-1)\xi(1)} & \cdots & \overline{\xi(N-1)\xi(N-1)} \end{bmatrix} \\
 &= \begin{bmatrix} R_\xi(0) & R_\xi(1) & R_\xi(2) & \cdots & R_\xi(N-1) \\ R_\xi(1) & R_\xi(0) & R_\xi(1) & \cdots & R_\xi(N-2) \\ R_\xi(2) & R_\xi(1) & R_\xi(0) & \cdots & R_\xi(N-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R_\xi(N-1) & R_\xi(N-2) & R_\xi(N-3) & \cdots & R_\xi(0) \end{bmatrix}.
 \end{aligned} \tag{8.9}$$

În mod absolut similar, se definește și matricea de autocovariație $\mathbf{K}_\xi$ ca fiind:

$$\mathbf{K}_\xi \triangleq \overline{(\boldsymbol{\xi} - \bar{\boldsymbol{\xi}})(\boldsymbol{\xi} - \bar{\boldsymbol{\xi}})^T} = \mathbf{R}_\xi - \bar{\boldsymbol{\xi}}\bar{\boldsymbol{\xi}}^T, \tag{8.10}$$

care, la fel ca și matricea de autocorelație, se poate scrie sub forma:

$$\mathbf{K}_\xi = \begin{bmatrix} K_\xi(0) & K_\xi(1) & K_\xi(2) & \cdots & K_\xi(N-1) \\ K_\xi(1) & K_\xi(0) & K_\xi(1) & \cdots & K_\xi(N-2) \\ K_\xi(2) & K_\xi(1) & K_\xi(0) & \cdots & K_\xi(N-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ K_\xi(N-1) & K_\xi(N-2) & K_\xi(N-3) & \cdots & K_\xi(0) \end{bmatrix}. \tag{8.11}$$

Se observă imediat din relațiile (8.9) și (8.11) că atât matricea de autocorelație cât și cea de autocovariație sunt simetrice și au o structură de tip Toeplitz¹.

O altă proprietate importantă a celor două matrici este aceea că ele sunt pozitiv definite. Vom demonstra această afirmație pentru matricea de autocorelație, demonstrația pentru matricea de autocovariație făcându-se în mod similar. Vom demonstra, deci, că $\forall \mathbf{x} \in \mathbb{R}_{N \times 1}$, avem

$$\mathbf{x}^T \mathbf{R}_\xi \mathbf{x} \geq 0. \tag{8.12}$$

Într-adevăr, fie un vector coloană oarecare $\mathbf{x} \in \mathbb{R}_{N \times 1}$ și fie variabila aleatoare scalară $\eta = \mathbf{x}^T \boldsymbol{\xi}$. Avem:

$$\overline{\eta^2} = \overline{\eta\eta^T} = \overline{\mathbf{x}^T \boldsymbol{\xi} (\mathbf{x}^T \boldsymbol{\xi})^T} = \overline{\mathbf{x}^T \boldsymbol{\xi} \boldsymbol{\xi}^T \mathbf{x}} = \mathbf{x}^T \overline{\boldsymbol{\xi} \boldsymbol{\xi}^T} \mathbf{x} = \mathbf{x}^T \mathbf{R}_\xi \mathbf{x}. \tag{8.13}$$

¹O matrice este de tip Toeplitz dacă elementele de pe orice diagonală paralelă cu diagonala principală sunt egale.

În demonstrația de mai sus am folosit faptul că η este un scalar, deci el este egal cu transpusa sa. Cum, prin definiție, $\overline{\eta^2} \geq 0$, rezultă proprietatea enunțată.

Faptul că matricea de autocorelație (autocovariație) este simetrică și pozitiv definită este echivalent cu a spune că valorile sale proprii sunt toate pozitive, proprietate la care vom mai face referire în cele ce urmează.

8.3 Modele stochastice

Termenul de model, aici, se referă la ipotezele ce se fac privind modul în care a fost generat un anumit semnal de interes. Din multitudinea de modele posibile, cele liniare prezintă un interes deosebit, datorită simplității lor. Problema care se pune, deci, este de a proiecta un sistem liniar invariant în timp (ceea ce am numit pe scurt *filtru liniar*) care să genereze un semnal cu proprietăți statistice cunoscute (dorite) pornind de la un semnal având un model statistic simplu, cum ar fi, de exemplu, un zgomot alb în timp discret², adică un semnal $\nu(n)$ de medie nulă, având funcția de autocorelație dată de:

$$R_\nu(n-k) = \overline{\nu(n)\nu(k)} = \begin{cases} \sigma_\nu^2 & \text{dacă } k = n \\ 0 & \text{dacă } k \neq n \end{cases}. \quad (8.14)$$

Problema este ilustrată în figura 8.2.

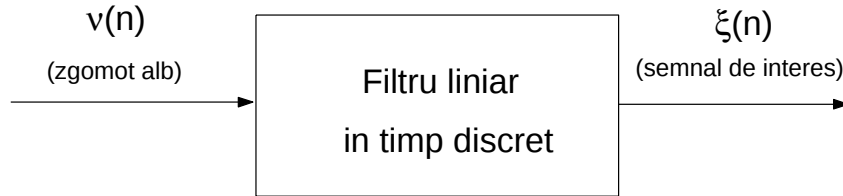


Figura 8.2: Model stohastic liniar.

Să precizăm că problema pe care ne propunem să o rezolvăm nu este nouă. În paragraful 5.11 s-a discutat, în contextul prelucrării semnalelor aleatoare continue, despre trecerea semnalelor aleatoare prin filtre liniare, deducându-se o relație de calcul al mărimilor statistice ale semnalului de ieșire (densitate spectrală de putere, sau, alternativ, funcție de autocorelație) în funcție de cele ale semnalului de intrare și de parametrii sistemului (presupuși cunoscuți). În cazul de față, avem aceeași schemă generală (un filtru liniar și două semnale, de intrare și ieșire) numai că problema pe care ne-o punem este de a determina parametrii filtrului care să conducă la obținerea unui semnal de parametri doriți, în condițiile în care modelul semnalului de intrare este unul cunoscut.

Există trei tipuri de modele liniare în timp discret, pe care le vom prezenta pe scurt în paragrafele următoare.

²Un astfel de semnal, mai exact o secvență de numere decorelate, este foarte ușor de generat cu calculatorul, prin apelul funcțiilor de generare de numere pseudo-aleatoare.

8.3.1 Modelul AR

Modelul auto-regresiv (AR) este caracterizat de următoarea relație între intrare și ieșire:

$$\xi(n) + a_1\xi(n-1) + a_2\xi(n-2) + \dots + a_M\xi(n-M) = \nu(n) \quad n = 0, 1, 2, \dots \quad (8.15)$$

unde constantele a_i se numesc coeficienții filtrului. Scriind relația (8.15) sub forma

$$\xi(n) = \nu(n) - \sum_{i=1}^M a_i \xi(n-i), \quad (8.16)$$

se observă în calculul eșantionului curent al semnalului de ieșire $\xi(n)$ intervin M eșantioane anterioare ale aceluiași semnal și eșantionul curent al semnalului de intrare, ceea ce justifică denumirea modelului. Schema bloc a unui model AR este prezentată în figura 8.3.a.

Dacă se consideră transformatele $\mathcal{Z}$ ale semnalelor, adică $N(z) = \mathcal{Z}\{\nu(n)\}(z)$, $X(z) = \mathcal{Z}\{\xi(n)\}(z)$, și a secvenței de coeficienți:

$$H_A(z) = \mathcal{Z}\{a_i\}(z) = \sum_{i=0}^M a_i z^{-i}, \quad (8.17)$$

unde $a_0 \stackrel{\text{not}}{=} 1$, atunci relația (8.15) se scrie în domeniul spectral ca:

$$H_A(z)X(z) = N(z), \quad (8.18)$$

de unde rezultă că funcția de transfer a filtrului AR este dată de:

$$H(z) = \frac{X(z)}{N(z)} = \frac{1}{H_A(z)} = \frac{1}{\sum_{i=0}^M a_i z^{-i}}. \quad (8.19)$$

Pentru filtrele care implică recursie, se pune în mod special problema stabilității. Un filtru se numește stabil dacă, în condițiile în care semnalul de intrare este mărginit, atunci și semnalul de ieșire este de asemenea mărginit:

$$\exists M \in \mathbb{R} \text{ astfel încât } |\nu(n)| \leq M \Rightarrow |\xi(n)| \leq M, \quad \forall n \in \mathbb{Z}. \quad (8.20)$$

După cum se știe din teoria filtrelor digitale, condiția necesară și suficientă ca un filtru să fie mărginit este ca polii funcției de transfer să fie în interiorul cercului unitate. Pentru cazul de față, notând cu p_i rădăcinile numitorului funcției de transfer:

$$H_A(p_i) = 0, \quad i = 1, \dots, M, \quad (8.21)$$

atunci condiția de stabilitate este echivalentă cu:

$$|p_i| < 1. \quad (8.22)$$

După cum este demonstrat în referința [4], condiția (8.22) asigură și staționaritatea (asimptotică) a semnalului de ieșire.

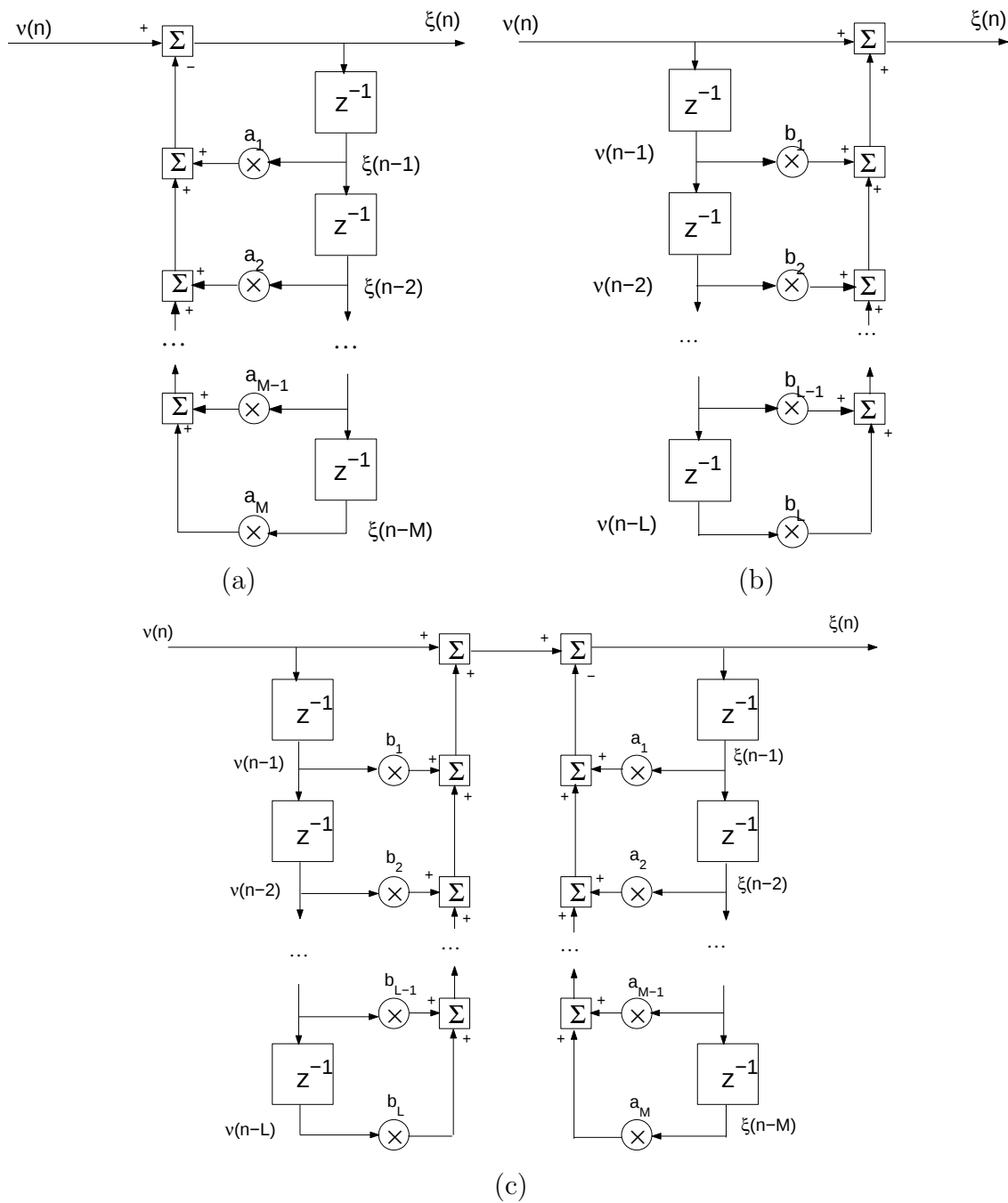


Figura 8.3: Schemele bloc ale diverselor tipuri de modele stochastice: (a) AR, (b) MA, (c) ARMA. Blocurile figurate cu z^{-1} sunt blocuri de întârziere cu un ciclu de ceas.

8.3.2 Modelul MA

Modelul “moving average” (MA)³ este caracterizat de relația:

$$\xi(n) = \nu(n) + b_1\nu(n-1) + b_2\nu(n-2) + \dots + b_L\nu(n-L) \quad n = 0, 1, 2, \dots \quad (8.23)$$

Cu alte cuvinte, eșantionul curent al semnalului de ieșire se calculează ca o medie ponderată a ultimelor $L + 1$ eșantioane ale semnalului de intrare cu coeficienți constanți, ceea ce justifică denumirea dată modelului. Schema bloc a unui filtru MA este desenată în figura 8.3.b.

Considerând transformata $\mathcal{Z}$ a secvenței de coeficienți $\{b_i\}$:

$$H_B(z) = \mathcal{Z}\{b_i\} = \sum_{i=0}^L b_i z^{-i}, \quad (8.24)$$

unde $b_0 \stackrel{\text{not}}{=} 1$, relația (8.23) se scrie în domeniul spectral ca:

$$X(z) = H_B(z)N(z), \quad (8.25)$$

de unde rezultă că funcția de transfer a filtrului MA este dată de:

$$H(z) = \frac{X(z)}{N(z)} = H_B(z) = \sum_{i=0}^L b_i z^{-i}. \quad (8.26)$$

Se observă că funcția de transfer a filtrului nu are numitor, deci filtrul este în mod inherent stabil (cum, dealtfel, se poate observa direct din relația ce leagă intrarea de ieșire în timp).

8.3.3 Modelul ARMA

Modelul ARMA este o combinație între cele două modele precedente, fiind caracterizat de relația:

$$\begin{aligned} \xi(n) + a_1\xi(n-1) + a_2\xi(n-2) + \dots + a_M\xi(n-M) = \\ \nu(n) + b_1\nu(n-1) + b_2\nu(n-2) + \dots + b_L\nu(n-L) \quad n = 0, 1, 2, \dots \end{aligned} \quad (8.27)$$

În calculul eșantionului curent al semnalului de ieșire intervin eșantioane anterioare atât ale acestuia, cât și ale semnalului de intrare. Schema bloc a unui filtru ARMA este prezentată în figura 8.3.c.

Cu notațiile din (8.17) și (8.24), relația (8.27) se scrie în domeniu spectral ca:

$$X(z)H_A(z) = N(z)H_B(z), \quad (8.28)$$

de unde rezultă că funcția de transfer a filtrului este dată de:

$$H(z) = \frac{X(z)}{N(z)} = \frac{H_B(z)}{H_A(z)}. \quad (8.29)$$

Problema stabilității este aceeași ca la modelul AR, și anume pentru ca filtrul ARMA să fie stabil trebuie să aibă loc (8.22), cu p_i dați de (8.21).

³Datorită faptului că termenul MA este consacrat în literatura de specialitate, vom păstra terminologia engleză, în locul celei românești, care s-ar putea traduce ca “medie glisantă”.

8.3.4 Ecuațiile Yule–Walker

Dintre cele trei modele descrise anterior, vom rezolva problema enunțată, și anume de proiectare a filtrului (respectiv de calcul al coeficienților acestuia) pentru a obține un semnal cu parametri statistici cunoscuți pentru un model de tip AR, întrucât permite exprimarea analitică a legăturii între coeficienți și parametrii statistici ai semnalului de ieșire.

Vom considera, deci, în continuare, că relația dintre semnalul de intrare și cel de ieșire este dată de (8.15). Înmulțind-o pe aceasta cu $\xi(n-i)$ și mediind statistic relația, obținem:

$$\overline{\xi(n)\xi(n-i)} + a_1\overline{\xi(n-1)\xi(n-i)} + \dots + a_M\overline{\xi(n-M)\xi(n-i)} = \overline{\nu(n)\xi(n-i)}. \quad (8.30)$$

Se observă că mediile statistice din partea stângă a ecuației reprezintă valori ale funcției de autocorelație a semnalului de ieșire. Mai mult, pentru $i > 0$ termenul din dreapta este nul, întrucât reprezintă corelația între eșantionul curent al zgomotului de la intrare și un eșantion anterior al semnalului de ieșire:

$$\overline{\nu(n)\xi(n-i)} = 0 \quad i = 1, 2, 3, \dots \quad (8.31)$$

Relația de mai sus este evidentă: $\xi(n-i)$ depinde de eșantioanele zgomotului de până la momentul $n-i$, care sunt toate decorelate cu cel de la momentul n . Astfel, relația (8.30) se scrie:

$$R_\xi(i) + a_1R_\xi(i-1) + a_2R_\xi(i-2) + \dots + a_MR_\xi(i-M) = 0, \quad i = 1, 2, 3, \dots \quad (8.32)$$

Scriind (8.32) pentru $i = 1, \dots, M$ obținem un sistem de M ecuații cu M necunoscute, care leagă coeficienții necunoscuți ai filtrului a_i de valorile cunoscute (dorite) ale funcției de autocorelație a semnalului de ieșire, sistem care poartă numele de ecuațiile Yule–Walker:

$$\begin{aligned} R_\xi(1) + a_1R_\xi(0) + a_2R_\xi(-1) + \dots + a_MR_\xi(-M+1) &= 0 \\ R_\xi(2) + a_1R_\xi(1) + a_2R_\xi(0) + \dots + a_MR_\xi(-M+2) &= 0 \\ \dots & \\ R_\xi(M) + a_1R_\xi(M-1) + a_2R_\xi(M-2) + \dots + a_MR_\xi(0) &= 0 \end{aligned} \quad (8.33)$$

Ținând cont de faptul că funcția de autocorelație R_ξ este pară, sistemul (8.33) poate fi scris compact, sub formă matricială ca:

$$\begin{bmatrix} R_\xi(0) & R_\xi(1) & R_\xi(2) & \dots & R_\xi(M-1) \\ R_\xi(1) & R_\xi(0) & R_\xi(1) & \dots & R_\xi(M-2) \\ R_\xi(2) & R_\xi(1) & R_\xi(0) & \dots & R_\xi(M-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R_\xi(M-1) & R_\xi(M-2) & R_\xi(M-3) & \dots & R_\xi(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_M \end{bmatrix} = - \begin{bmatrix} R_\xi(1) \\ R_\xi(2) \\ R_\xi(3) \\ \vdots \\ R_\xi(M) \end{bmatrix}. \quad (8.34)$$

Observând că matricea sistemului de ecuații este chiar matricea de autocorelație a semnalului de ieșire (8.9), și făcând următoarele notații suplimentare:

$$\mathbf{a} \stackrel{\text{not}}{=} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_M \end{bmatrix}, \quad (8.35)$$

respectiv

$$\mathbf{r}_\xi \stackrel{\text{not}}{=} \begin{bmatrix} R_\xi(1) \\ R_\xi(2) \\ R_\xi(3) \\ \vdots \\ R_\xi(M) \end{bmatrix}, \quad (8.36)$$

soluția sistemului de ecuații Yule–Walker se scrie ca:

$$\mathbf{a} = -\mathbf{R}_\xi^{-1} \mathbf{r}_\xi. \quad (8.37)$$

Se impun câteva observații în legătură cu rezultatul obținut mai sus. Să pornim prin a scrie relația (8.30) pentru $i = 0$. În acest caz, termenul din dreapta nu mai este zero, ci, ținând cont de (8.16) poate fi scris:

$$\overline{\nu(n)\xi(n)} = \overline{\nu(n) \left(\nu(n) - \sum_{i=1}^M a_i \xi(n-i) \right)} = \underbrace{\overline{\nu^2(n)}}_{\sigma_\nu^2} - \sum_{i=1}^M a_i \underbrace{\overline{\nu(n)\xi(n-i)}}_0 = \sigma_\nu^2, \quad (8.38)$$

de unde (8.30) se scrie:

$$R_\xi(0) + a_1 R_\xi(1) + a_2 R_\xi(2) + \dots + a_M R_\xi(M) = \sigma_\nu^2. \quad (8.39)$$

Această relație ne dă modul de calcul al varianței zgomotului, care, în condițiile modelului luat în considerare (zgomot alb), este singurul parametru necunoscut al acestuia. Practic, sistemul de ecuații (8.33) împreună cu ecuația (8.39) formează împreună un sistem care leagă în mod biunivoc $M + 1$ necunoscute (respectiv cei M coeficienți a_i ai filtrului și varianța zgomotului de intrare σ_ν^2) de primele $M + 1$ valori ale funcției de autocorelație a semnalului de ieșire $R_\xi(i)$ cu $i = 0, \dots, M$. Se observă că dacă împărțim (8.34) la $R_\xi(0)$, obținem un sistem de ecuații care leagă valorile funcției de autocorelație normalizate

$$\rho_\xi(i) = \frac{R_\xi(i)}{R_\xi(0)} \quad (8.40)$$

de coeficienții filtrului:

$$\{a_1, a_2, \dots, a_M\} \Leftrightarrow \{\rho_\xi(1), \rho_\xi(2), \dots, \rho_\xi(M)\}. \quad (8.41)$$

Cu alte cuvinte, coeficienții a_i determină forma generală a funcției de autocorelație, din varianța zgomotului σ_v^2 reglând, apoi, valorile propriu-zise (nenormalizate) ale acesteia:

$$\sigma_v^2 \Leftrightarrow R_\xi(0). \quad (8.42)$$

O altă observație importantă este aceea că relația (8.32) poate fi scrisă și pentru $i > M$, adică:

$$\begin{aligned} R_\xi(M+1) + a_1 R_\xi(M) + a_2 R_\xi(M-1) + \dots + a_M R_\xi(1) &= 0 \\ R_\xi(M+2) + a_1 R_\xi(M+1) + a_2 R_\xi(M) + \dots + a_M R_\xi(2) &= 0 \\ \dots \end{aligned} \quad (8.43)$$

Aceste relații permit calculul valorilor funcției de autocorelație $R_\xi(i)$ pentru $i > M$ în funcție de $R_\xi(0), R_\xi(1), \dots, R_\xi(M)$. Deci, valorile funcției de autocorelație pentru $i > M$ nu sunt independente, ci depind de primele $M+1$ valori, respectiv cele specificate. Această observație este utilă pentru alegerea ordinului modelului (respectiv a numărului M de coeficienți ai filtrului), întrucât acesta reprezintă numărul de “grade de libertate”, adică numărul de valori în care putem specifica independent funcția de autocorelație pe care dorim să o obținem la ieșire.

În figura 8.4 sunt prezentate două forme dorite ale funcției de autocorelație, iar în figura 8.5 sunt prezentate semnalele obținute la ieșirea filtrului AR corespunzătoare funcțiilor de autocorelație specificate.

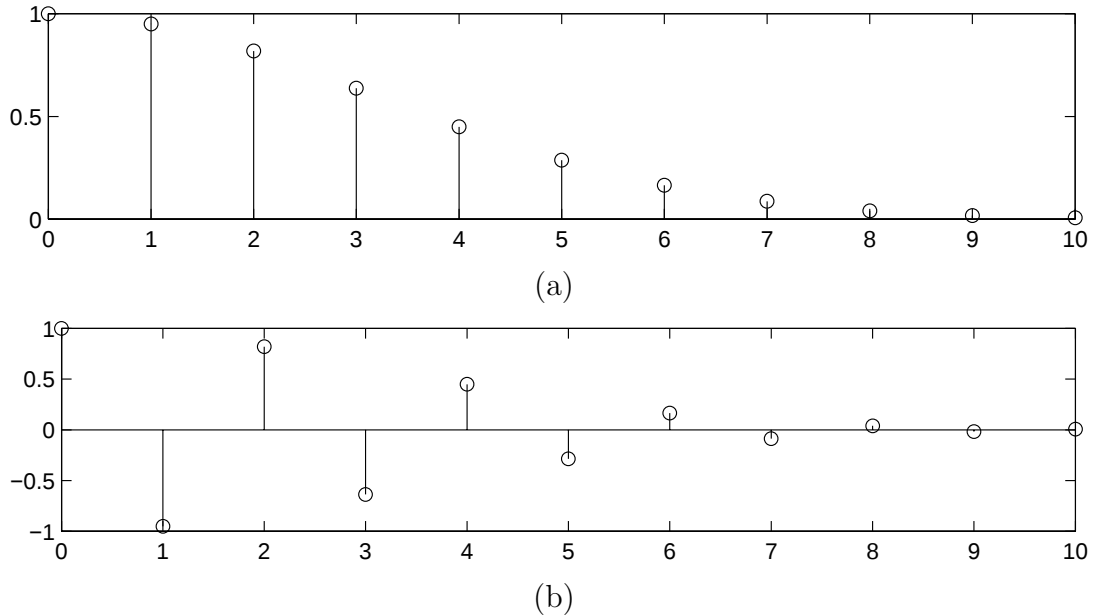


Figura 8.4: Două exemple de funcție de autocorelație $R_\xi(i)$ dorită la ieșirea unui filtru AR, specificate până la ordinul $M = 10$.

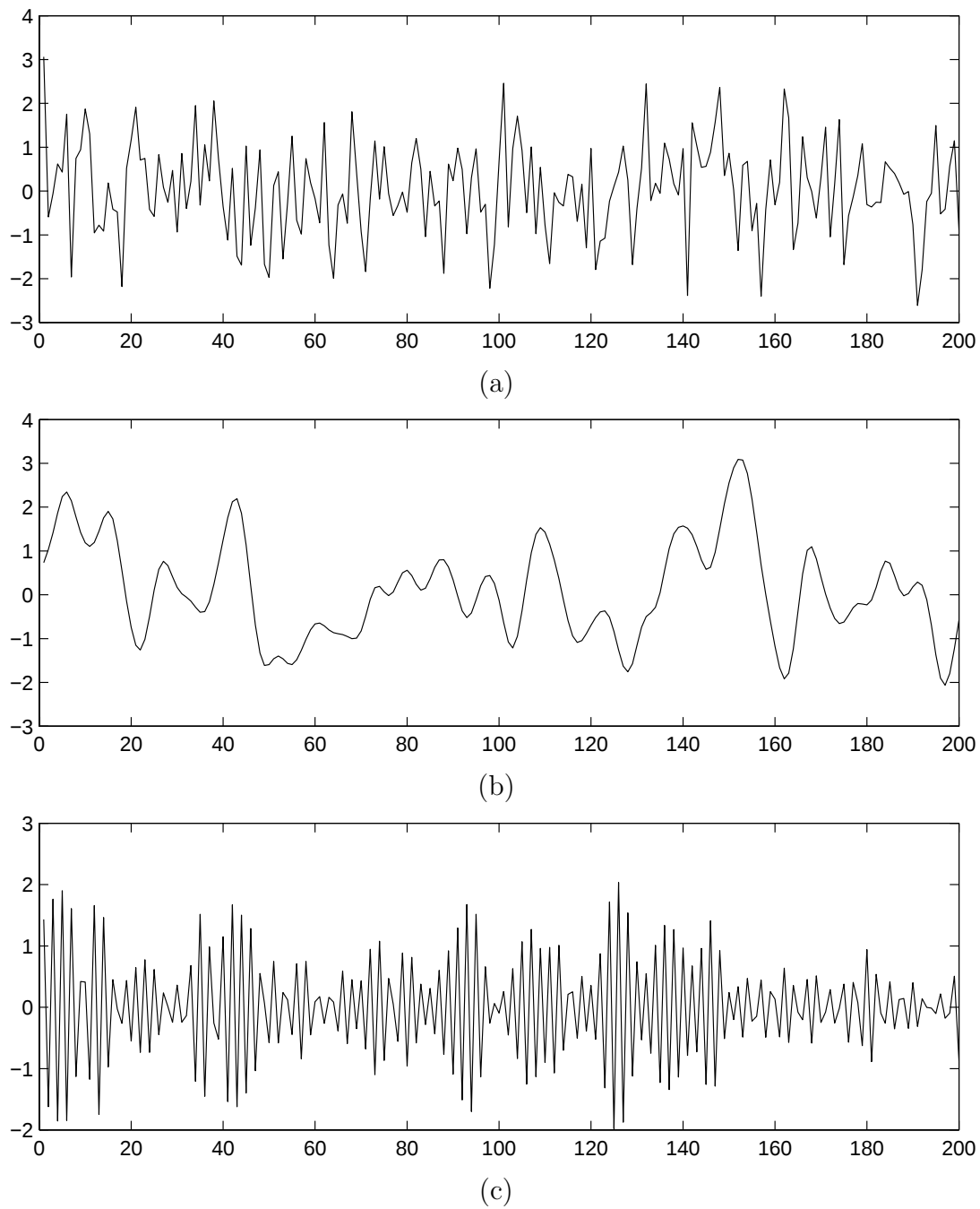


Figura 8.5: Semnalele la intrarea și ieșirea modelului AR: (a) Semnalul de intrare (zgomot alb); (b) Semnalul la ieșirea filtrului AR de ordin $M = 10$ proiectat în funcție de valorile $R_\xi(i)$ din figura 8.4.a; (c) Semnalul la ieșirea filtrului AR proiectat $M = 10$ în funcție de valorile $R_\xi(i)$ din figura 8.4.b

Capitolul 9

Filtrarea optimală a semnalelor

9.1 Introducere

Problema filtrării optimale a semnalelor, pe care o vom trata în acest capitol pentru semnale aleatoare în timp discret, se poate enunța astfel: dorim să proiectăm un filtru liniar care, aplicându-i-se la intrare un semnal aleator $\xi(n)$, să producă la ieșire un semnal $\eta(n)$ care să fie cât mai apropiat de un semnal aleator dorit $\gamma(n)$. Facem din start precizarea că atât semnalul de intrare, cât și cel dorit a fi obținut la ieșire, sunt presupuse staționare și de medie nulă. Putem exprima similaritatea între semnalul dorit și cel obținut la ieșirea filtrului prin intermediul diferenței între cele două semnale:

$$e(n) = \gamma(n) - \eta(n) \quad n = 0, 1, 2, \dots, \quad (9.1)$$

numit semnal eroare. Evident, este de dorit ca $e(n)$ să aibă valori cât mai mici. Problema este ilustrată în figura 9.1.

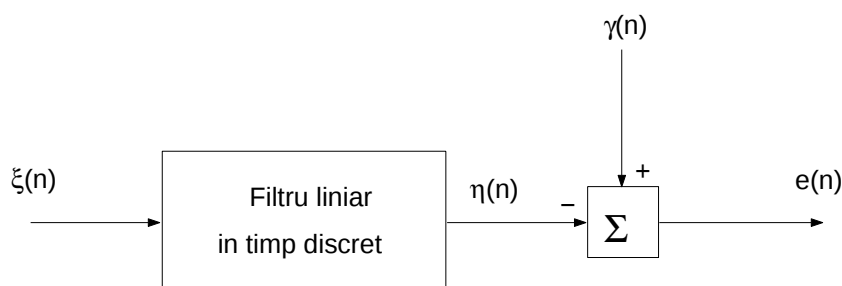


Figura 9.1: Punerea problemei filtrării optimale a semnalelor.

Datorită constrângerii pe care o impunem din start sistemului căutat, de a fi liniar și invariant în timp, relația dintre semnalele de intrare și ieșire este dată de:

$$\eta(n) = \sum_{k=0}^{\infty} h_k \xi(n-k). \quad (9.2)$$

Relația (9.2) nu reprezintă altceva decât adaptarea relației (5.89) pentru cazul semnalelor în timp discret, considerând în plus filtrul cauzal. Valorile h_k se numesc coeficienții filtrului și constituie necunoscutele problemei puse. Sunt posibile două tipuri de structuri de filtre: cele cu număr finit de coeficienți, care se numesc filtre de tip FIR (*Finite Impulse Response*), și cele cu un număr infinit de coeficienți, care se numesc filtre IIR (*Infinite Impulse Response*)¹.

Pentru a putea pune problema în ecuație, mai trebuie definit un criteriu de eroare, prin care să traducem matematic dezideratul enunțat, și anume asemănarea semnalului de ieșire $\eta(n)$ cu cel dorit $\gamma(n)$. Alegem, ca de altfel în majoritatea cazurilor tratate în această lucrare, minimizarea erorii pătratice medii, criteriu de eroare care conduce, în general, la un sistem de ecuații liniare, deci ușor rezolvabil. Ținând cont de (9.2), semnalul eroare se scrie ca:

$$e(n) = \gamma(n) - \sum_{k=0}^{\infty} h_k \xi(n-k) \quad n = 0, 1, 2, \dots \quad (9.3)$$

iar eroarea medie pătratică ε se definește ca:

$$\varepsilon = \overline{e(n)^2}, \quad (9.4)$$

care, datorită staționarității semnalelor, este independentă de momentul de timp ales n .

Astfel, problema enunțată se traduce matematic prin găsirea ponderilor h_k care minimizează pe ε dat de (9.4). Filtrul optimal astfel obținut se numește filtru Wiener.

9.2 Principiul ortogonalității

Valorile ponderilor h_i care minimizează eroarea pătratică medie sunt acele valori în care derivata acesteia este nulă. Derivata lui ε după ponderea h_i se scrie ca:

$$\frac{\partial \varepsilon}{\partial h_i} = \frac{\partial \overline{e^2(n)}}{\partial h_i} = \overline{2e(n) \frac{\partial e(n)}{\partial h_i}} = -2\overline{e(n)\xi(n-i)}, \quad \forall i = 0, 1, 2, \dots, \quad (9.5)$$

de unde rezultă că:

$$\overline{e_o(n)\xi(n-i)} = 0, \quad \forall i = 0, 1, 2, \dots, \quad (9.6)$$

unde cu $e_o(n)$ am notat valoarea optimă a erorii, adică cea corespunzătoare lui $\varepsilon_{\min}$. Relația (9.6) este expresia *principiului ortogonalității*, și, deși constituie un rezultat intermediar, are o interpretare importantă. Principiul ortogonalității afirmă că, în cazul optimal, diferența între ce dorim și ce avem la ieșirea filtrului la un moment de timp dat este decorelată cu toate eșantioanele anterioare ale semnalului de intrare! Această decorelare a erorii față de semnalul de intrare semnifică faptul că am extras maximum de informație din acesta, întrucât nu mai există nici o legătură statistică între semnalul de care dispunem (cel de intrare) și ceea ce mai lipsește semnalului de ieșire pentru ca acesta să corespundă întru totul cerințelor noastre (adică să fie egal cu semnalul dorit $\gamma(n)$).

¹Din punct de vedere al implementării practice, filtrele IIR sunt filtre cu recursie, de tipul AR sau ARMA, în timp ce filtrele de FIR sunt filtre de tip MA.

Corelația între semnalul de ieșire și eroarea optimală poate fi scrisă, ținând cont de (9.2), ca:

$$\overline{e_o(n)\eta(n-i)} = \overline{e_o(n) \left(\sum_{k=0}^{\infty} h_{k,o} \xi(n-i-k) \right)} = \sum_{k=0}^{\infty} h_{k,o} \overline{e_o(n) \xi(n-i-k)}, \quad (9.7)$$

unde cu $h_{k,o}$ am notat ponderile optimale, care minimizează pe ε . Înlocuind (9.6) în (9.7), rezultă că

$$\overline{e_o(n)\eta(n-i)} = 0, \quad \forall i = 0, 1, 2, \dots, \quad (9.8)$$

relație ce reprezintă corolarul principiului ortogonalității, care afirmă că eroarea obținută în cazul optimal este decorelată și cu valorile prezentă și anterioare ale semnalului de ieșire.

9.3 Ecuatiile Wiener–Hopf

În acest paragraf, vom deduce relațiile de calcul al ponderilor filtrului Wiener. Pornind de la principiul ortogonalității (9.6), și înlocuind expresia erorii (9.3) în cazul optimal, obținem:

$$\overline{\left(\gamma(n) - \sum_{k=0}^{\infty} h_{k,o} \xi(n-k) \right) \xi(n-i)} = 0, \quad \forall i = 0, 1, 2, \dots \quad (9.9)$$

de unde, prin calcule elementare, obținem:

$$\underbrace{\overline{\gamma(n)\xi(n-i)}}_{R_{\xi\gamma}(i)} = \sum_{k=0}^{\infty} h_{k,o} \underbrace{\overline{\xi(n-i)\xi(n-k)}}_{R_{\xi}(i-k)} \quad \forall i = 0, 1, 2, \dots \quad (9.10)$$

Prin identificarea mediilor din ecuația de mai sus cu valorile funcției de autocorelație a semnalului de intrare R_{ξ} , respectiv a funcției de intercorelație între semnalul de intrare și cel dorit la ieșirea filtrului $R_{\xi\gamma}$, obținem relația căutată:

$$\sum_{k=0}^{\infty} h_{k,o} R_{\xi}(i-k) = R_{\xi\gamma}(i), \quad \forall i = 0, 1, 2, 3, \dots, \quad (9.11)$$

Ecuatiile (9.11) poartă numele de ecuațiile Wiener–Hopf, și reprezintă un sistem liniar de o infinitate de ecuații cu o infinitate de necunoscute, care permite calculul ponderilor optimale $h_{k,o}$ în funcție de autocorelația semnalului de intrare și de intercorelația dintre semnalul de intrare și cel dorit la ieșirea filtrului (presupuse cunoscute).

Să facem următoarea observație importantă. Soluția sistemului de ecuații Wiener–Hopf poate fi scrisă în domeniul spectral, observând că termenul stâng al relației (9.11) nu reprezintă altceva decât convoluția în timp discret între funcția de autocorelație a

semnalului de intrare și secvența $\{h_{k,o}\}_{k \in \mathbb{N}}$ a ponderilor filtrului Wiener. Rezultă, deci, că relația (9.11) poate fi scrisă în domeniul spectral sub forma:

$$q_\xi(\omega)H_o(\omega) = q_{\xi\gamma}(\omega), \quad \forall \omega \in [-\pi, \pi], \quad (9.12)$$

unde $H_o(\omega)$ este funcția de transfer a filtrului optimal, adică transformata Fourier în timp discret a secvenței de coeficienți $h_{k,o}$:

$$H_o(\omega) = \mathcal{F}\{h_{k,o}\}(\omega) = \sum_{k=0}^{\infty} h_{k,o} \exp(-j\omega k) \quad (9.13)$$

Ecuția (9.12) permite calculul funcției de transfer a filtrului optimal ca:

$$H_o(\omega) = \frac{q_{\xi\gamma}(\omega)}{q_\xi(\omega)}. \quad (9.14)$$

În paragraful următor, vom da o soluție a sistemului de ecuații Wiener–Hopf în domeniul temporal, pentru o structură particulară de filtru, și anume filtru cu număr finit de coeficienți.

9.3.1 Ecuațiile Wiener–Hopf pentru filtru de tip FIR

Considerând numărul de coeficienți ai filtrului finit, adică:

$$h_k = 0 \quad \forall k = M, M+1, M+2, \dots, \quad (9.15)$$

relația dintre ieșire și intrarea filtrului este dată de:

$$\eta(n) = \sum_{k=0}^{M-1} h_k \xi(n-k). \quad (9.16)$$

Structura filtrului Wiener de tip FIR, care implementează relația (9.16) este prezentată în figura 9.2.

În acest caz particular, sistemul de ecuații Wiener–Hopf (9.11) devine:

$$\sum_{k=0}^{M-1} h_{k,o} R_\xi(i-k) = R_{\xi\gamma}(i), \quad \forall i = 0, 1, 2, 3, \dots \quad (9.17)$$

Scriind relațiile de mai sus pentru $i = 0, 1, \dots, M-1$, obținem următorul sistem de M ecuații cu M necunoscute, a cărui soluție vor fi ponderile filtrului optimal $h_{k,o}$:

$$\begin{aligned} h_{0,o}R_\xi(0) + h_{1,o}R_\xi(-1) + \dots + h_{M-1,o}R_\xi(-M+1) &= R_{\xi\gamma}(0) \\ h_{0,o}R_\xi(1) + h_{1,o}R_\xi(0) + \dots + h_{M-1,o}R_\xi(-M+2) &= R_{\xi\gamma}(1) \\ \dots & \\ h_{0,o}R_\xi(M-1) + h_{1,o}R_\xi(M-2) + \dots + h_{M-1,o}R_\xi(0) &= R_{\xi\gamma}(M-1) \end{aligned} \quad (9.18)$$

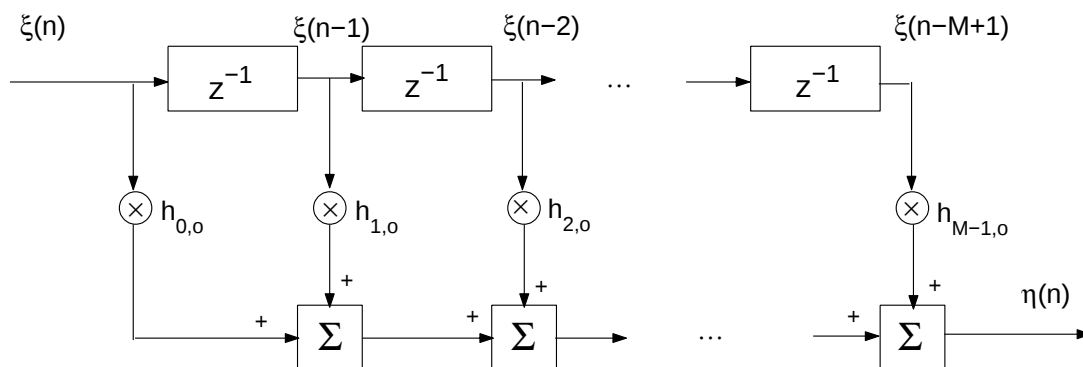


Figura 9.2: Schema bloc a unui filtru optimal cu număr finit de coeficienți (filtru de tip FIR).

Ținând cont de faptul că funcția de autocorelație R_ξ este pară, sistemul (9.18) poate fi scris sub formă matricială ca:

$$\underbrace{\begin{bmatrix} R_\xi(0) & R_\xi(1) & \cdots & R_\xi(M-1) \\ R_\xi(1) & R_\xi(0) & \cdots & R_\xi(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ R_\xi(M-1) & R_\xi(M-2) & \cdots & R_\xi(0) \end{bmatrix}}_{\mathbf{R}_\xi} \begin{bmatrix} h_{0,o} \\ h_{1,o} \\ \vdots \\ h_{M-1,o} \end{bmatrix} = \begin{bmatrix} R_{\xi\gamma}(0) \\ R_{\xi\gamma}(1) \\ \vdots \\ R_{\xi\gamma}(M-1) \end{bmatrix}. \quad (9.19)$$

Notând cu $\mathbf{h}_o$ vectorul coloană al ponderilor optimale :

$$\mathbf{h}_o \stackrel{\text{not}}{=} \begin{bmatrix} h_{0,o} \\ h_{1,o} \\ \vdots \\ h_{M-1,o} \end{bmatrix}, \quad (9.20)$$

și cu $\mathbf{r}_{\xi\gamma}$ vectorul ce conține valorile funcției de intercorelație între semnalul de intrare și cel dorit la ieșire:

$$\mathbf{r}_{\xi\gamma} \stackrel{\text{not}}{=} \begin{bmatrix} R_{\xi\gamma}(0) \\ R_{\xi\gamma}(1) \\ \vdots \\ R_{\xi\gamma}(M-1) \end{bmatrix}, \quad (9.21)$$

putem scrie soluția sistemului Wiener-Hopf pentru cazul unui filtru de tip FIR ca:

$$\mathbf{h}_o = \mathbf{R}_\xi^{-1} \mathbf{r}_{\xi\gamma}. \quad (9.22)$$

9.4 Aplicații ale filtrării optimale a semnalelor

În paragrafele următoare vom prezenta două dintre cele mai importante aplicații ale filtrelor optimale, și anume atenuarea zgomotului și predicția.

9.4.1 Atenuarea zgomotului

Să presupunem că semnalul de intrare $\xi(n)$ reprezintă versiunea perturbată cu zgomot aditiv $\nu(n)$ a unui semnal de interes $\zeta(n)$

$$\xi(n) = \zeta(n) + \nu(n) \quad n = 0, 1, 2, \dots \quad (9.23)$$

În acest caz, dorim ca ieșirea filtrului să aproximeze cât mai bine semnalul util. Cu alte cuvinte, alegem:

$$\gamma(n) \equiv \zeta(n). \quad (9.24)$$

Să presupunem în continuare că semnalul $\nu(n)$ reprezintă un proces aleator zgomot alb, independent de semnalul util, de unde rezultă că funcția de autocorelație a semnalului de intrare este:

$$\begin{aligned} R_\xi(k) &= \overline{\xi(n)\xi(n+k)} = \overline{(\zeta(n) + \nu(n))(\zeta(n+k) + \nu(n+k))} \\ &= \underbrace{\overline{\zeta(n)\zeta(n+k)}}_{R_\zeta(k)} + \underbrace{\overline{\zeta(n)\nu(n+k)}}_0 + \underbrace{\overline{\zeta(n+k)\nu(n)}}_0 + \underbrace{\overline{\nu(n)\nu(n+k)}}_{R_\nu(k)} \\ &= R_\zeta(k) + R_\nu(k). \end{aligned} \quad (9.25)$$

În dezvoltarea de mai sus am folosit faptul că procesele ζ și ν sunt independente, ceea ce face ca $\overline{\zeta(n_1)\nu(n_2)} = \overline{\zeta(n_1)} \underbrace{\overline{\nu(n_2)}}_0 = 0$.

În plus, intercorelația între ξ și γ se scrie:

$$R_{\xi\gamma}(i) = \overline{\xi(n)\gamma(n+i)} = \overline{(\zeta(n) + \nu(n))\zeta(n+i)} = \underbrace{\overline{\zeta(n)\zeta(n+i)}}_{R_\zeta(i)} + \underbrace{\overline{\nu(n)\zeta(n+i)}}_0 = R_\zeta(i), \quad (9.26)$$

de unde rezultă că vectorul $\mathbf{r}_{\xi\gamma}$ este prima coloană a matricii de autocorelație a semnalului util $\mathbf{R}_\zeta$:

$$\mathbf{r}_{\xi\gamma} = \begin{bmatrix} R_\zeta(0) \\ R_\zeta(1) \\ \vdots \\ R_\zeta(M-1) \end{bmatrix} \stackrel{\text{not}}{=} \mathbf{r}_\zeta \quad (9.27)$$

Ținând cont de (8.14), rezultă că matricea de autocorelație a zgomotului este diagonală: $\mathbf{R}_\nu = \sigma_\nu^2 \mathbf{I}_N$. Prin particularizarea relației (9.22) pentru datele deduse mai sus, rezultă forma finală a coeficienților filtrului Wiener optimal pentru atenuarea zgomotului:

$$\mathbf{h}_o = (\mathbf{R}_\zeta + \sigma_\nu^2 \mathbf{I}_N)^{-1} \mathbf{r}_\zeta. \quad (9.28)$$

În figura 9.3 este prezentat un exemplu de filtrare de zgomot cu un filtru Wiener.

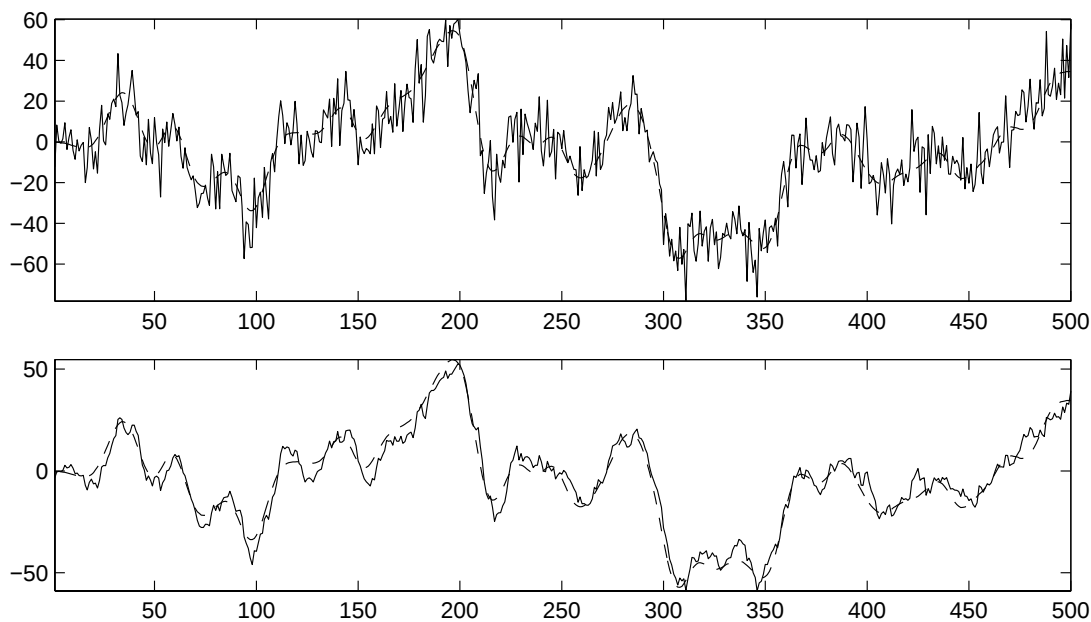


Figura 9.3: Exemplu de atenuare de zgomot cu un filtru optimal: în figura de sus, semnalul original (cu linie întreruptă) și cel perturbat cu zgomot alb (cu linie continuă). În figura de jos, semnalul original (cu linie întreruptă) și semnalul rezultat prin filtrarea cu un filtru Wiener cu 20 de coeficienți (cu linie continuă).

9.4.2 Predicția

Predicția este o altă aplicație importantă în prelucrarea semnalelor. Scopul predicției este ca la fiecare moment de timp n să se estimeze valoarea următoare a semnalului de intrare.

Astfel, considerând semnalul de intrare $\xi(n)$, alegem semnalul de ieșire dorit al filtrului Wiener predictor ca fiind:

$$\gamma(n) \equiv \xi(n+1). \quad (9.29)$$

În aceste condiții, intercorelația între semnalul de intrare și cel dorit este:

$$R_{\xi\gamma}(i) = \overline{\xi(n)\gamma(n+i)} = \overline{\xi(n)\xi(n+i+1)} = R_{\xi}(i+1), \quad (9.30)$$

ceea ce face ca vectorul $\mathbf{r}_{\xi\gamma}$ din (9.21) să fie:

$$\mathbf{r}_{\xi\gamma} = \begin{bmatrix} R_{\xi}(1) \\ R_{\xi}(2) \\ \vdots \\ R_{\xi}(M) \end{bmatrix} \stackrel{\text{not}}{=} \mathbf{r}_{\xi}. \quad (9.31)$$

Vectorul ponderilor optimale ale filtrului Wiener predictor este astfel dat de:

$$\mathbf{h}_o = \mathbf{R}_{\xi}^{-1} \mathbf{r}_{\xi}. \quad (9.32)$$

Se impun două observații referitoare la relația (9.32). Prima este legată de faptul că problema predicției unui semnal se pune numai în cazul în care între eșantioanele semnalului respectiv există o oarecare corelație. Într-adevăr, dacă $\xi(n)$ este complet decorelat, adică $R_\xi(i) = 0$ pentru $i \neq 0$, atunci vectorul $\mathbf{r}_\xi$ din (9.31) este nul, și la fel va fi și vectorul ponderilor optimale $\mathbf{h}_o$ și, implicit, și ieșirea filtrului predictor! Se poate de asemenea arăta că eroarea de predicție este cu atât mai mică cu cât între eșantioanele semnalului există o corelație mai puternică.

A doua observație se referă la similaritatea între soluția sistemului de ecuații Wiener–Hopf pentru filtrul predictor (9.32) și soluția sistemului de ecuații Yule–Walker (8.37). Această asemănare nu este întâmplătoare: dacă privim filtrul Wiener predictor din figura 9.1 ca un sistem cu intrarea dată de $\xi(n)$ și cu ieșirea dată de eroarea de predicție $e(n) = \hat{\xi}(n+1) - \xi(n)$ (în loc de valoarea prezisă $\eta(n) \stackrel{\text{not}}{=} \hat{\xi}(n+1)$), atunci acesta poate fi văzut ca inversul filtrului auto-regresiv, respectiv un sistem liniar care are la intrare un semnal cu funcție de autocorelație cunoscută și care trebuie să producă la ieșire un semnal complet decorelat cu intrarea.

În figura 9.4 este prezentat un exemplu de acțiune a unui filtru Wiener predictor.

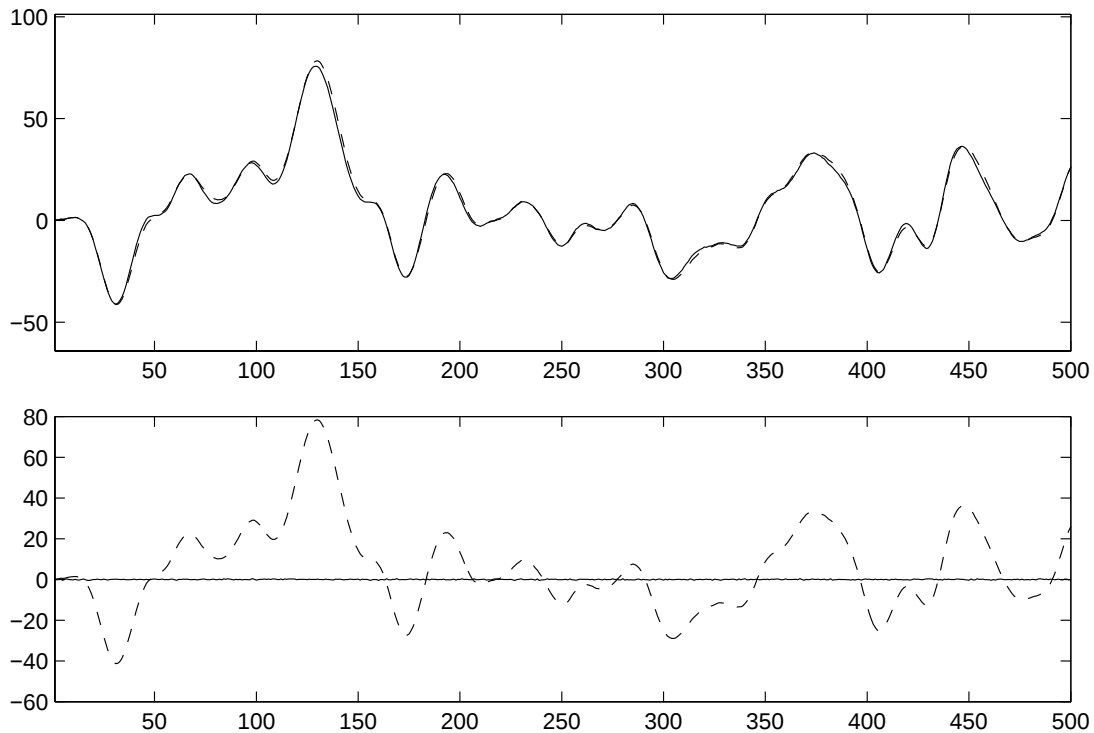


Figura 9.4: Exemplu de predicție: în figura de sus, semnalul original (cu linie întreruptă) și cel “prezis” de un filtru Wiener cu 20 de coeficienți (cu linie continuă). În figura de jos, sunt reprezentate semnalul original (cu linie întreruptă) și eroarea de predicție (cu linie continuă).

Predicția este utilizată pentru compresia semnalelor (spre exemplu, semnalul vocal se

codează foarte bine prin predicție liniară). Dacă predicția este bună, deci dacă valoarea “prezisă” aproximează fidel valoarea reală a semnalului, atunci pe canal se poate transmite numai diferența între cele două valori, care, fiind în medie mică (vezi figura 9.4), poate fi codată pe un număr redus de biți. Având în vedere că regula de predicție este cunoscută, partea predictibilă a semnalului poate fi reprodusă exact la recepție. Din aceasta, prin adăugarea erorii transmise pe canal, poate fi reconstruit fiecare eșantion al semnalului. Astfel, se poate face o transmisiune cu compresie a semnalului, practic fără pierderi, pe baza redundanței (corelației) existente în acesta.

Schema bloc a unui lanț de transmisiune cu predicție, conform celor descrise mai sus, este prezentată în figura 9.5. Blocul $\mathbf{Q}$ este bloc de codare a erorii: într-o primă fază, eroarea este cuantizată iar apoi este codată cu un cod entropic (de tip Huffman) ce exploatează distribuția puternic neuniformă a valorilor acesteia: valorile mici au probabilitate de apariție mare, în timp ce valorile mari apar rar. Blocul $\mathbf{Q}^{-1}$ reprezintă bloc de decodare a erorii. Singurele pierderi (diferența între $e(n)$ și $e'(n)$) sunt cele datorate operației de cuantizare².

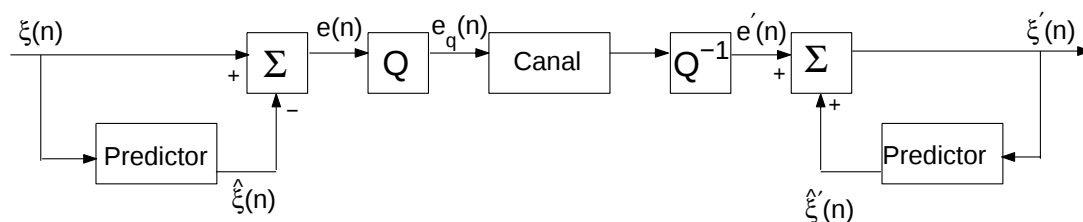


Figura 9.5: Schema bloc a unui lanț de transmisiune cu predicție.

²Problema cuantizării semnalelor este tratată în detaliu în capitolul 11.

Capitolul 10

Transformate unitare

10.1 Introducere

Termenul de tranformată unitară se referă la o clasă de matrici *unitare* folosite pentru reprezentarea semnalelor. O matrice pătratică având componente complexe $\mathbf{A} \in \mathbb{C}_{N \times N}$ se numește unitară dacă:

$$\mathbf{A}^{-1} = \mathbf{A}^{*T}, \quad (10.1)$$

unde prin $()^*$ am notat operatorul de conjugare complexă.

Să menționăm chiar de la început că, deși cazul complex nu este deloc neglijabil¹, în acest capitol vom trata numai problema transformatelor cu coeficienți reali, deci vom presupune în continuare $\mathbf{A} \in \mathbb{R}_{N \times N}$, caz în care condiția de unitaritate devine:

$$\mathbf{A}^{-1} = \mathbf{A}^T. \quad (10.2)$$

Fiind dată matricea $\mathbf{A}$ unitară și vectorul coloană $\boldsymbol{\xi} = [\xi(0) \ \xi(1) \ \cdots \ \xi(N-1)]^T$, se definește transformata lui $\boldsymbol{\xi}$ ca fiind vectorul coloană $\boldsymbol{\eta} = [\eta(0) \ \eta(1) \ \cdots \ \eta(N-1)]^T$ obținut ca:

$$\boldsymbol{\eta} = \mathbf{A}\boldsymbol{\xi}. \quad (10.3)$$

Considerând elementele matricii $\mathbf{A} = \{a(k, n)\}_{k,n=0,\dots,N-1}$, relația (10.3) se poate scrie:

$$\eta(k) = \sum_{n=0}^{N-1} a(k, n)\xi(n), \quad \forall k = 0, 1, \dots, N-1. \quad (10.4)$$

Transformata inversă, care permite recuperarea vectorului original $\boldsymbol{\xi}$ din cel transformat $\boldsymbol{\eta}$ se scrie, ținând cont de (10.2):

$$\boldsymbol{\xi} = \mathbf{A}^T \boldsymbol{\eta}, \quad (10.5)$$

¹Spre exemplu, transformata Fourier discretă se obține cu ajutorul unei matrici $\mathbf{A}$ având elemente complexe!

sau, element cu element:

$$\xi(n) = \sum_{k=0}^{N-1} a(k, n) \eta(k), \quad \forall n = 0, 1, \dots, N-1. \quad (10.6)$$

Semnificația unei transformate unitare poate fi înțeleasă scriind matricea $\mathbf{A}^T$ sub forma

$$\mathbf{A}^T = \begin{bmatrix} \mathbf{a}_0 & \mathbf{a}_1 & \cdots & \mathbf{a}_{N-1} \end{bmatrix}, \quad (10.7)$$

unde $\mathbf{a}_k$ sunt vectori coloană:

$$\mathbf{a}_k = \begin{bmatrix} a(k, 0) \\ a(k, 1) \\ \vdots \\ a(k, N-1) \end{bmatrix}. \quad (10.8)$$

Cu aceste notații, relațiile care dau transformata directă (10.4), respectiv pe cea inversă (10.5), pot fi scrise sub forma:

$$\eta(k) = \boldsymbol{\xi}^T \mathbf{a}_k = \langle \boldsymbol{\xi}, \mathbf{a}_k \rangle \quad \forall k = 0, 1, \dots, N-1, \quad (10.9)$$

respectiv:

$$\boldsymbol{\xi} = \sum_{k=0}^{N-1} \eta(k) \mathbf{a}_k, \quad (10.10)$$

unde cu $\langle \mathbf{f}, \mathbf{g} \rangle$ a fost notat produsul scalar între vectorii coloană $\mathbf{f}, \mathbf{g} \in \mathbb{R}_{N \times 1}$, definit ca:

$$\langle \mathbf{f}, \mathbf{g} \rangle \triangleq \mathbf{f}^T \mathbf{g} = \sum_{i=0}^{N-1} f(i)g(i). \quad (10.11)$$

Calculul lui $\boldsymbol{\xi}$ din relația (10.10) se interpretează ca o *descompunere* a acestuia într-o altă bază a lui $\mathbb{R}^N$, și anume cea formată din coloanele matricii $\mathbf{A}^T$: $\{\mathbf{a}_0, \mathbf{a}_1, \dots, \mathbf{a}_{N-1}\}$. Coeficienții dezvoltării lui $\boldsymbol{\xi}$ în această bază sunt chiar componentele $\eta(k)$ ale vectorului transformat $\boldsymbol{\eta}$. Conform (10.9), aceștia se obțin prin produs scalar între vectorul dezvoltat $\boldsymbol{\xi}$ și axa $\mathbf{a}_k$, care nu reprezintă altceva decât *proiecția* lui $\boldsymbol{\xi}$ pe $\mathbf{a}_k$.

În plus, noua bază în care se reprezintă semnalul este ortonormală, fapt asigurat de condiția (10.2) de unitaritate impusă matricii $\mathbf{A}$. Într-adevăr, (10.2) se poate rescrie ca:

$$\mathbf{A} \mathbf{A}^T = \mathbf{I}_N, \quad (10.12)$$

ceea ce, scris explicit în funcție de vectorii $\mathbf{a}_k$, devine:

$$\begin{bmatrix} \mathbf{a}_0^T \\ \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_{N-1}^T \end{bmatrix} \begin{bmatrix} \mathbf{a}_0 & \mathbf{a}_1 & \cdots & \mathbf{a}_{N-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}. \quad (10.13)$$

Din relația de mai sus, rezultă că:

$$\langle \mathbf{a}_i, \mathbf{a}_j \rangle = \mathbf{a}_i^T \mathbf{a}_j = \begin{cases} 1 & \text{dacă } i = j \\ 0 & \text{în rest} \end{cases}, \quad (10.14)$$

ceea ce este echivalent cu a spune că baza $\{\mathbf{a}_k\}_{k=0,\dots,N-1}$ pe care se descompune vectorul original $\boldsymbol{\xi}$ este ortonormală, întrucât fiecare doi vectori din set sunt perpendiculari, iar norma fiecăruia este unu!

Pentru fixarea noțiunilor, să considerăm un exemplu pentru care avem o reprezentare geometrică, respectiv $N = 2$. Fie vectorul $\boldsymbol{\xi} = [5, 2]^T$ din figura 10.1. Cele două componente ale vectorului $\boldsymbol{\xi}$ reprezintă coeficienții dezvoltării vectorului în baza canonică $\{\mathbf{e}_x = [1, 0]^T, \mathbf{e}_y = [0, 1]^T\}$ aleasă din oficiu pentru reprezentarea oricărui vector din $\mathbb{R}^2$. Cu alte cuvinte, $\boldsymbol{\xi} = [5, 2]^T$ este echivalent cu:

$$\boldsymbol{\xi} = 5\mathbf{e}_x + 2\mathbf{e}_y,$$

cu

$$\langle \boldsymbol{\xi}, \mathbf{e}_x \rangle = 5$$

$$\langle \boldsymbol{\xi}, \mathbf{e}_y \rangle = 2$$

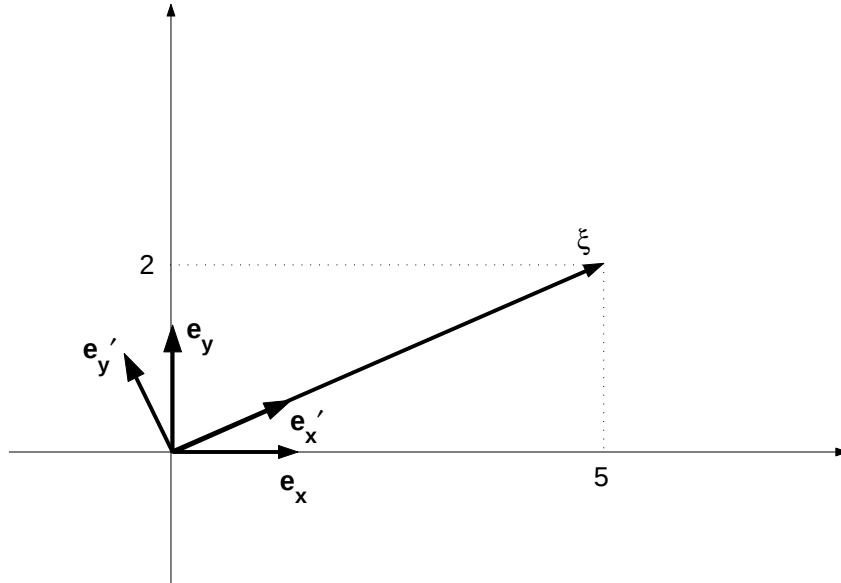


Figura 10.1: Exemplu de schimbare de bază.

Însă $\{\mathbf{e}_x, \mathbf{e}_y\}$ nu este singura bază a lui $\mathbb{R}^2$. Să considerăm, spre exemplu, baza $\{\mathbf{e}'_x, \mathbf{e}'_y\}$ obținută printr-o rotație a bazei canonice cu un unghi astfel ales încât versorul $\mathbf{e}'_x$ să devină coliniar cu vectorul considerat $\boldsymbol{\xi}$. În noua bază $\{\mathbf{e}'_x, \mathbf{e}'_y\}$, componentele lui $\boldsymbol{\xi}$ devin:

$$\langle \boldsymbol{\xi}, \mathbf{e}'_x \rangle = \sqrt{29}$$

$$\langle \boldsymbol{\xi}, \mathbf{e}'_y \rangle = 0$$

iar acesta poate fi scris ca:

$$\boldsymbol{\xi} = \sqrt{29}\mathbf{e}'_x + 0\mathbf{e}'_y = \underbrace{\begin{bmatrix} \mathbf{e}'_x & \mathbf{e}'_y \end{bmatrix}}_{\mathbf{A}^T} \underbrace{\begin{bmatrix} \sqrt{29} \\ 0 \end{bmatrix}}_{\boldsymbol{\eta}}.$$

Deci, vectorul transformat $\boldsymbol{\eta} = [\sqrt{29}, 0]^T$ este noua reprezentare a lui $\boldsymbol{\xi}$ în baza considerată.

Dintre aplicațiile transformatelor unitare, cea mai importantă este compresia de date, care este posibilă datorită proprietății transformatelor utilizate în practică de a *compacta* energia într-un număr redus de coeficienți în spațiul transformatei. Să precizăm pentru început că o transformată unitară conservă energia semnalului. Într-adevăr:

$$E_{\boldsymbol{\eta}} = \|\boldsymbol{\eta}\|^2 = \boldsymbol{\eta}^T \boldsymbol{\eta} = (\mathbf{A}\boldsymbol{\xi})^T \mathbf{A}\boldsymbol{\xi} = \boldsymbol{\xi}^T \underbrace{\mathbf{A}^T \mathbf{A}}_{\mathbf{I}_N} \boldsymbol{\xi} = \boldsymbol{\xi}^T \boldsymbol{\xi} = \|\boldsymbol{\xi}\|^2 = E_{\boldsymbol{\xi}}. \quad (10.15)$$

Deși energia totală se conservă, pentru majoritatea transformatelor utilizate în practică, aceasta tinde să fie distribuită inegal între coeficienții spațiului transformat. Cu alte cuvinte, energia este concentrată într-un număr redus de coeficienți ai transformării, restul putând fi neglijăți, pierderea de informație indusă fiind extrem de mică.

Problema compresiei optimale cu transformate unitare este tratată în continuarea acestui capitol.

10.2 Transformata Karhunen–Loève

Problema care se pune este aceea de calcul al matricii unitare

$$\mathbf{L} = \begin{bmatrix} \mathbf{l}_0 & \mathbf{l}_1 & \cdots & \mathbf{l}_{N-1} \end{bmatrix}^T \quad (10.16)$$

cu ajutorul căreia să compactăm cât mai bine energia în spațiul transformatei, cu alte cuvinte, să concentrăm energia într-un număr minim de coeficienți în spațiul transformării.

Fie $\boldsymbol{\eta}$ transformata vectorului original $\boldsymbol{\xi}$:

$$\boldsymbol{\eta} = \mathbf{L}\boldsymbol{\xi} = \begin{bmatrix} \eta(0) & \cdots & \eta(m-1) & \eta(m) & \cdots & \eta(N-1) \end{bmatrix}^T. \quad (10.17)$$

Din vectorul $\boldsymbol{\eta}$ păstrăm numai primele m componente, restul înlocuindu-le cu niște constante c_i . Cu alte cuvinte, îl aproximăm pe $\boldsymbol{\eta}$ cu $\tilde{\boldsymbol{\eta}}$ dat de:

$$\tilde{\boldsymbol{\eta}} = \begin{bmatrix} \eta(0) & \cdots & \eta(m-1) & c_m & \cdots & c_{N-1} \end{bmatrix}^T. \quad (10.18)$$

Cu ajutorul vectorului aproximat în spațiul original

$$\tilde{\boldsymbol{\xi}} = \mathbf{L}^T \tilde{\boldsymbol{\eta}} \quad (10.19)$$

se definește eroarea pătratică medie indusă de aproximarea în spațiul transformatei ca fiind:

$$\varepsilon = \overline{\|\boldsymbol{\xi} - \tilde{\boldsymbol{\xi}}\|^2}. \quad (10.20)$$

Problema din punct de vedere matematic este de determinare a matricii transformării $\mathbf{L}$ astfel încât eroarea pătratică medie (10.20) să fie minimă.

Trecând eroarea pătratică medie din (10.20) în spațiul transformatei, avem:

$$\begin{aligned}\varepsilon &= \overline{(\boldsymbol{\xi} - \tilde{\boldsymbol{\xi}})^T (\boldsymbol{\xi} - \tilde{\boldsymbol{\xi}})} = \overline{(\mathbf{L}^T (\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}}))^T (\mathbf{L}^T (\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}}))} = \overline{(\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}})^T \underbrace{\mathbf{L} \mathbf{L}^T}_{\mathbf{I}_N} (\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}})} \\ &= \overline{(\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}})^T (\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}})} = \overline{\|\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}}\|^2} = \sum_{i=0}^{N-1} \overline{(\eta(i) - \tilde{\eta}(i))^2} = \sum_{i=0}^{N-1} \overline{(\eta(i) - \tilde{\eta}(i))^2}.\end{aligned}\quad (10.21)$$

Ținând cont că primele m componente ale lui $\boldsymbol{\eta}$ și $\tilde{\boldsymbol{\eta}}$ sunt identice, rezultă că eroarea pătratică medie poate fi scrisă ca:

$$\varepsilon = \sum_{i=m}^{N-1} \overline{(\eta(i) - c_i)^2}. \quad (10.22)$$

În acest punct, putem determina constantele c_k care minimizează eroarea ε . Facem, deci:

$$\frac{\partial \varepsilon}{\partial c_k} = \frac{\partial}{\partial c_k} \left(\sum_{i=m}^{N-1} \overline{(\eta(i) - c_i)^2} \right) = \overline{-2(\eta(k) - c_k)} = 0, \quad (10.23)$$

de unde rezultă:

$$c_k = \overline{\eta(k)} \quad \forall k = m, \dots, N-1. \quad (10.24)$$

Înlocuind (10.24) în (10.22), obținem:

$$\varepsilon = \sum_{i=m}^{N-1} \overline{(\eta(i) - \overline{\eta(i)})^2}. \quad (10.25)$$

Dar, conform (10.9), avem

$$\eta(i) = \boldsymbol{\xi}^T \mathbf{l}_i = \mathbf{l}_i^T \boldsymbol{\xi}, \quad (10.26)$$

de unde:

$$\begin{aligned}\varepsilon &= \sum_{i=m}^{N-1} \overline{(\eta(i) - \overline{\eta(i)}) (\eta(i) - \overline{\eta(i)})} = \sum_{i=m}^{N-1} \overline{(\mathbf{l}_i^T \boldsymbol{\xi} - \overline{\mathbf{l}_i^T \boldsymbol{\xi}}) (\boldsymbol{\xi}^T \mathbf{l}_i - \overline{\boldsymbol{\xi}^T \mathbf{l}_i})} \\ &= \sum_{i=m}^{N-1} \overline{\mathbf{l}_i^T (\boldsymbol{\xi} - \overline{\boldsymbol{\xi}}) (\boldsymbol{\xi} - \overline{\boldsymbol{\xi}})^T \mathbf{l}_i} = \sum_{i=m}^{N-1} \overline{\mathbf{l}_i^T \underbrace{(\boldsymbol{\xi} - \overline{\boldsymbol{\xi}}) (\boldsymbol{\xi} - \overline{\boldsymbol{\xi}})^T}_{\mathbf{K}_\xi} \mathbf{l}_i} = \sum_{i=m}^{N-1} \overline{\mathbf{l}_i^T \mathbf{K}_\xi \mathbf{l}_i}.\end{aligned}\quad (10.27)$$

Relația (10.27) ne dă scrierea cantității de minimizat ε în funcție de necunoscutele $\mathbf{l}_i$. Nu putem, însă, trece direct la minimizarea lui ε fără a ține cont de constrângerile asupra vectorilor $\mathbf{l}_i$ de a fi de normă unitară, întrucât sunt coloanele unei matrici unitare. Astfel, minimizarea lui ε cu constrângerile $\mathbf{l}_i^T \mathbf{l}_i = 1$ revine la minimizarea expresiei:

$$\Psi = \sum_{i=m}^{N-1} (\mathbf{l}_i^T \mathbf{K}_\xi \mathbf{l}_i - \lambda_i (\mathbf{l}_i^T \mathbf{l}_i - 1)), \quad (10.28)$$

unde $\lambda_i \in \mathbb{R}$ sunt multiplicatorii Lagrange “atașați” constrângerilor impuse.

Pentru rezolvarea elegantă, sub formă vectorială, a problemei minimizării lui Ψ din (10.28), definim gradientul vectorial al unei expresii scalare. Fie $E(\mathbf{v})$ o funcție scalară de vectorul $\mathbf{v} = [v_1, \dots, v_N]^T$. Definim gradientul lui E după $\mathbf{v}$ ca fiind:

$$\nabla_{\mathbf{v}} E = \begin{bmatrix} \frac{\partial E}{\partial v_1} \\ \vdots \\ \frac{\partial E}{\partial v_N} \end{bmatrix}. \quad (10.29)$$

Evident, vectorul $\mathbf{v}$ care minimizează pe E se obține făcând:

$$\nabla_{\mathbf{v}} E = \mathbf{0}. \quad (10.30)$$

Mai mult, se poate arăta prin calcul direct că pentru orice matrice simetrică ($\mathbf{B} = \mathbf{B}^T$) avem:

$$\nabla_{\mathbf{v}} (\mathbf{v}^T \mathbf{B} \mathbf{v}) = 2\mathbf{B} \mathbf{v}. \quad (10.31)$$

Ținând cont că matricea de covariație $\mathbf{K}_{\xi}$ este simetrică, și că putem scrie $\mathbf{l}_i^T \mathbf{l}_i = \mathbf{l}_i^T \mathbf{I}_N \mathbf{l}_i$, atunci, aplicând rezultatele enunțate mai sus, rezultă că vectorii $\mathbf{v}_k$ care minimizează pe Ψ se obțin din ecuația:

$$\nabla_{\mathbf{l}_k} \Psi = 2\mathbf{K}_{\xi} \mathbf{l}_k - 2\lambda_k \mathbf{l}_k = \mathbf{0}, \quad (10.32)$$

de unde rezultă că:

$$\mathbf{K}_{\xi} \mathbf{l}_k = \lambda_k \mathbf{l}_k. \quad (10.33)$$

Cu alte cuvinte, coloanele matricii transformatei Karhunen–Loève (prescurtată în cele ce urmează KL) $\mathbf{L}^T$, cu ajutorul căreia se compactează optim energia vectorului original ξ , sunt *vectorii proprii* ai matricii de covariație ai vectorului respectiv! De asemenea, multiplicatorii Lagrange λ_i sunt valorile proprii atașate vectorilor $\mathbf{l}_i$.

Să facem câteva observații importante referitoare la rezultatul obținut. În primul rând, înlocuind vectorii $\mathbf{l}_k$ din (10.33) în expresia (10.27), obținem valoarea minimă a erorii pătratice medii:

$$\varepsilon_{\min} = \sum_{i=m}^{N-1} \mathbf{l}_i^T \underbrace{\mathbf{K}_{\xi} \mathbf{l}_i}_{\lambda_i \mathbf{l}_i} = \sum_{i=m}^{N-1} \lambda_i \underbrace{\mathbf{l}_i^T \mathbf{l}_i}_1 = \sum_{i=m}^{N-1} \lambda_i. \quad (10.34)$$

Cu alte cuvinte, eroarea pătratică medie obținută în cazul optimal este suma valorilor proprii aferente axelor ale căror componente sunt “suprimate”. Din această observație deducem ordinea de așezare a coloanelor $\mathbf{l}_i$ în matricea $\mathbf{L}^T$: dacă dorim să eliminăm componentele vectorului $\boldsymbol{\eta}$ localizate la coadă (așa cum am impus la început) atunci vectorii $\mathbf{l}_k$ trebuie ordonați în sens descrescător al valorilor proprii aferente:

$$\mathbf{L}^T = \begin{bmatrix} \mathbf{l}_0 & \mathbf{l}_1 & \cdots & \mathbf{l}_{N-1} \end{bmatrix} \quad (10.35)$$

$$\lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{N-1}$$

În continuare, calculând matricea de covarianță a vectorului transformat $\boldsymbol{\eta}$, avem:

$$\begin{aligned} \mathbf{K}_{\boldsymbol{\eta}} &= \overline{(\boldsymbol{\eta} - \bar{\boldsymbol{\eta}})(\boldsymbol{\eta} - \bar{\boldsymbol{\eta}})^T} = \overline{\mathbf{L}(\boldsymbol{\xi} - \bar{\boldsymbol{\xi}})(\mathbf{L}(\boldsymbol{\xi} - \bar{\boldsymbol{\xi}}))^T} = \overline{\mathbf{L}(\boldsymbol{\xi} - \bar{\boldsymbol{\xi}})(\boldsymbol{\xi} - \bar{\boldsymbol{\xi}})^T \mathbf{L}^T} \\ &= \overline{\mathbf{L}(\boldsymbol{\xi} - \bar{\boldsymbol{\xi}})(\boldsymbol{\xi} - \bar{\boldsymbol{\xi}})^T} \mathbf{L}^T = \mathbf{L} \mathbf{K}_{\boldsymbol{\xi}} \mathbf{L}^T. \end{aligned} \quad (10.36)$$

Ținând cont de (10.33), ecuația (10.36) se scrie mai departe:

$$\begin{aligned} \mathbf{K}_{\boldsymbol{\eta}} &= \underbrace{\begin{bmatrix} \mathbf{l}_0^T \\ \mathbf{l}_1^T \\ \vdots \\ \mathbf{l}_{N-1}^T \end{bmatrix}}_{\mathbf{L}} \underbrace{\mathbf{K}_{\boldsymbol{\xi}} \begin{bmatrix} \mathbf{l}_0 & \mathbf{l}_1 & \cdots & \mathbf{l}_{N-1} \end{bmatrix}}_{\mathbf{L}^T} = \begin{bmatrix} \mathbf{l}_0^T \\ \mathbf{l}_1^T \\ \vdots \\ \mathbf{l}_{N-1}^T \end{bmatrix} \begin{bmatrix} \mathbf{K}_{\boldsymbol{\xi}} \mathbf{l}_0 & \mathbf{K}_{\boldsymbol{\xi}} \mathbf{l}_1 & \cdots & \mathbf{K}_{\boldsymbol{\xi}} \mathbf{l}_{N-1} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{l}_0^T \\ \mathbf{l}_1^T \\ \vdots \\ \mathbf{l}_{N-1}^T \end{bmatrix} \begin{bmatrix} \lambda_0 \mathbf{l}_0 & \lambda_1 \mathbf{l}_1 & \cdots & \lambda_{N-1} \mathbf{l}_{N-1} \end{bmatrix} = \begin{bmatrix} \lambda_0 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_1 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_{N-1} \end{bmatrix}. \end{aligned} \quad (10.37)$$

Semnificația rezultatului (10.37) este aceea că, în noile coordonate, coeficienții dezvoltării sunt complet decorelați! Mai mult, se observă că

$$\lambda_i = \sigma_{\boldsymbol{\eta}(i)}^2, \quad (10.38)$$

adică valorile λ_i au semnificația de varianță a coeficientului descompunerii vectorului $\boldsymbol{\xi}$ pe axa $\mathbf{l}_i$, ceea ce justifică alegerea făcută mai înainte, de eliminare a componentelor pe axele cu λ_i mic. Într-adevăr, în lumina relației (10.38), aceasta semnifică înlocuirea cu media lor a componentelor de varianță mică, ceea ce este natural.

În figura 10.2, este prezentat un exemplu de schimbare de coordonate cu transformata KL pentru un set de vectori 3D cu componente puternic corelate (ceea ce se poate observa din forma liniară a norului de puncte reprezentând vectorii respectivi). Se observă că în noile coordonate, una din axe este chiar axa principală a norului de puncte, ceea ce este echivalent cu a spune că, în medie, componentele vectorilor pe celelalte două axe (perpendiculare pe axa principală) vor fi mici, ele putând fi eliminate cu o pierdere de informație redusă.

10.3 Aproximări practice ale transformatei KL

Transformata KL se folosește rareori ca atare în practică. Unul dintre motive este acela că transformata nu este aceeași pentru orice semnal, ci depinde de statistica semnalului respectiv. Ea trebuie, deci, calculată pentru fiecare clasă de semnale în parte. Un alt motiv, și mai important, este acela că transformata este lentă, complexitatea algoritmului de implementare fiind $\mathcal{O}(N^2)$. Astfel, în practică se folosesc diverse aproximări ale

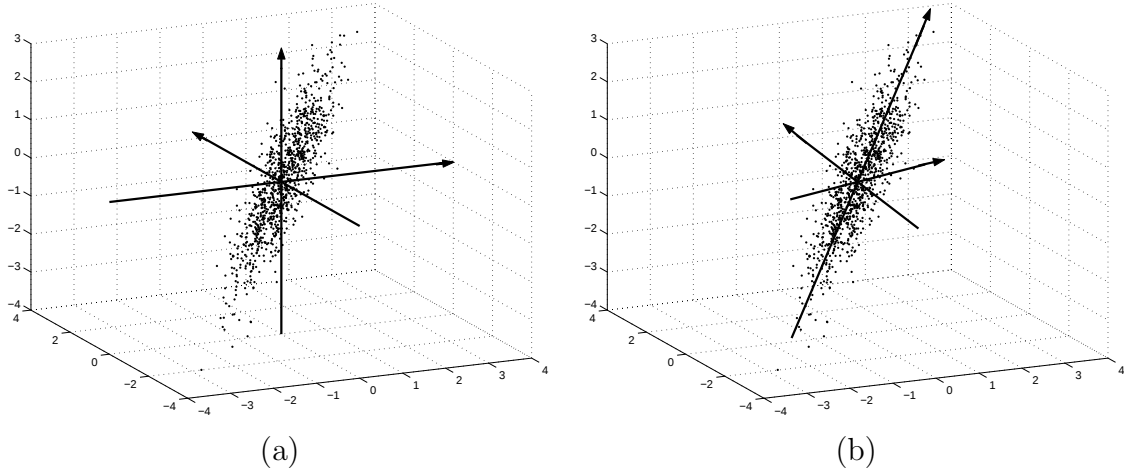


Figura 10.2: Exemplu de schimbare de coordonate a unui set de vectori cu componente puternic corelate. (a) Sistemul de coordonate canonic; (b) Sistemul de coordonate rezultat în urma aplicării transformatei KL.

transformatei, suboptimale, dar cu posibilități de implementare eficientă. În paragrafele următoare, vom prezenta două dintre transformate folosite în practică drept substitut de transformată KL.

10.3.1 Transformata cosinus discretă

Transformata cosinus discretă este definită de matricea unitară $\mathbf{C} = \{c(k, n)\}_{k,n=0,\dots,N-1}$ având elementele date de:

$$c(k, n) = \begin{cases} \frac{1}{\sqrt{N}} & \text{dacă } k = 0 \\ \sqrt{\frac{2}{N}} \cos \frac{\pi(2n+1)k}{2N} & \text{dacă } k \neq 0 \end{cases} \quad (10.39)$$

Se poate demonstra (vezi referința [5]) că transformata cosinus discretă poate fi implementată printr-un algoritm rapid, de complexitate² $\mathcal{O}(N \log N)$.

În ceea ce privește asemănarea transformatei cosinus discrete cu transformata optimă KL, vom arăta în continuare că matricea $\mathbf{C}$ aproximează foarte bine matricea transformatei KL pentru semnale Markov de ordinul unu puternic corelate.

Prin definiție, un semnal Markov de ordinul unu este un semnal aleator al cărui trecut nu are nici o influență asupra viitorului, cu condiția ca prezentul să fie specificat. Aceasta se traduce matematic ca:

$$w_\xi(x_n | x_{n-1}, x_{n-2}, x_{n-3} \dots) = w_\xi(x_n | x_{n-1}) \quad (10.40)$$

unde cu $w_\xi(x_n | x_{n-1}, x_{n-2}, x_{n-3} \dots)$ am notat densitatea de probabilitate a lui $\xi(n)$

²Pentru $N = 1000$, reducerea complexității de la $\mathcal{O}(N^2)$ la $\mathcal{O}(N \log N)$ este echivalentă cu creșterea vitezei de calcul de aproximativ 100 de ori!

condiționată de faptul că valorile anterioare ale semnalului au fost $x_{n-1}, x_{n-2}, \dots$:

$$w_\xi(x_n | x_{n-1}, x_{n-2}, \dots) \stackrel{\text{not}}{=} w_{\xi(n)}(\{\{\xi(n-1)=x_{n-1}\} \cap \{\xi(n-2)=x_{n-2}\} \cap \dots\})(x_n). \quad (10.41)$$

Relația (10.40) se interpretează astfel: pentru determinarea statisticii semnalului la momentul următor $n+1$ este suficientă cunoașterea valorii semnalului la momentul prezent n^3 . Semnalele Markov servesc deseori de model statistic simplificat pentru semnale reale.

Se poate arăta că funcția de autocovariație a unui semnal Markov de ordinul unu este de forma:

$$K_\xi(n) = \sigma_\xi^2 \rho^{|n|} \quad \text{cu } |\rho| < 1, \quad (10.42)$$

ceea ce este echivalent cu a spune că matricea de autocorelație a procesului este:

$$\mathbf{K}_\xi = \sigma_\xi^2 \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{N-1} \\ \rho & 1 & \rho & \dots & \rho^{N-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \dots & 1 \end{bmatrix}. \quad (10.43)$$

Prin calcul direct, se poate demonstra că inversa matricii de covariație din (10.43) este dată de:

$$\mathbf{K}_\xi^{-1} = \beta \begin{bmatrix} 1 - \rho\alpha & -\alpha & 0 & 0 & \dots & 0 & 0 & 0 \\ -\alpha & 1 & -\alpha & 0 & \dots & 0 & 0 & 0 \\ 0 & -\alpha & 1 & -\alpha & \dots & 0 & 0 & 0 \\ 0 & 0 & -\alpha & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & -\alpha & 0 \\ 0 & 0 & 0 & 0 & \dots & -\alpha & 1 & -\alpha \\ 0 & 0 & 0 & 0 & \dots & 0 & -\alpha & 1 - \rho\alpha \end{bmatrix}, \quad (10.44)$$

cu $\beta = \frac{1}{\sigma_\xi^2} \frac{1+\rho^2}{1-\rho^2}$ și $\alpha = \frac{\rho}{1+\rho^2}$.

Pe de altă parte, se poate demonstra prin calcul direct că vectorii coloană ai matricii $\mathbf{C}^T$ sunt vectorii proprii ai matricii:

$$\mathbf{M} = \begin{bmatrix} 1 - \alpha & -\alpha & 0 & 0 & \dots & 0 & 0 & 0 \\ -\alpha & 1 & -\alpha & 0 & \dots & 0 & 0 & 0 \\ 0 & -\alpha & 1 & -\alpha & \dots & 0 & 0 & 0 \\ 0 & 0 & -\alpha & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & -\alpha & 0 \\ 0 & 0 & 0 & 0 & \dots & -\alpha & 1 & -\alpha \\ 0 & 0 & 0 & 0 & \dots & 0 & -\alpha & 1 - \alpha \end{bmatrix}, \quad (10.45)$$

³Un semnal Markov de ordinul unu poate fi modelat ca fiind ieșirea unui model AR de ordinul unu: $\xi(n) = -a_1 \xi(n-1) + v(n)$.

cu $\alpha \in \mathbb{R}$ oarecare. Cu alte cuvinte, are loc relația:

$$\mathbf{M}\mathbf{c}_k = \lambda_k \mathbf{c}_k, \quad k = 0, \dots, N-1 \quad (10.46)$$

unde:

$$\mathbf{c}_k = [c(k, 0) \ c(k, 1) \ \dots \ c(k, N-1)]^T. \quad (10.47)$$

Prin compararea relației (10.44) cu (10.45), se observă că:

$$\mathbf{K}_\xi^{-1} \underset{|\rho| \lesssim 1}{\approx} \mathbf{M}. \quad (10.48)$$

Altfel spus, ținând cont de (10.46) și de (10.48), vectorii coloană ai matricii $\mathbf{C}^T$ aproximează cu bună precizie vectorii proprii ai matricii $\mathbf{K}_\xi^{-1}$ (care sunt identici cu cei ai lui $\mathbf{K}_\xi^4$) atunci când $\rho \lesssim 1$.

Concluzia este cea enunțată anterior: transformata cosinus discretă aproximează foarte bine transformata optimală KL pentru semnale puternic corelate. În practică, transformata cosinus discretă este folosită în standardul de compresie de imagini JPEG (după cum vom prezenta în detaliu în capitolul 12). În figura 10.3 este prezentat un exemplu de compresie cu transformata cosinus discretă.

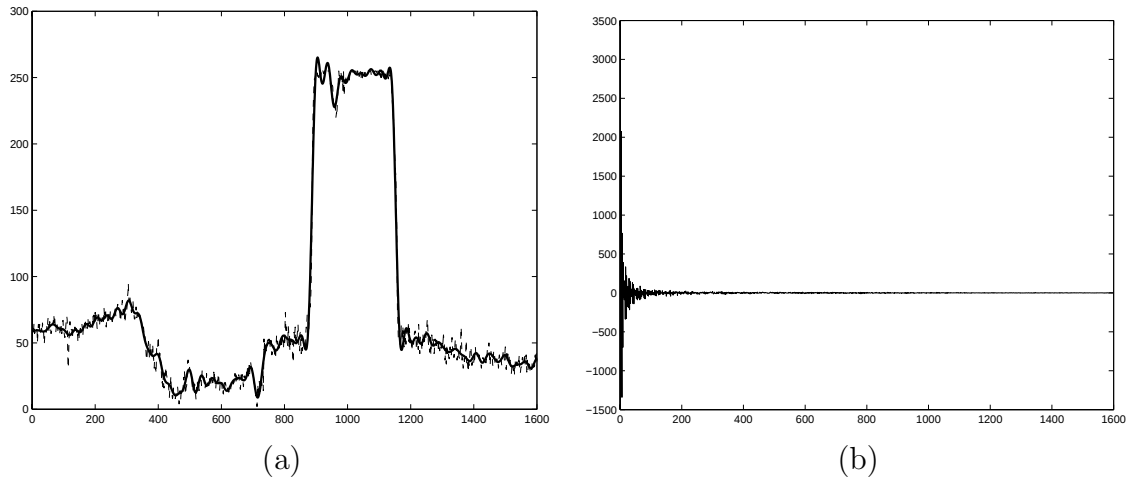


Figura 10.3: Exemplu de compresie de semnale cu transformata cosinus discretă: (a) Semnalul original (cu linie punctată) și cel rezultat în urma eliminării a 87,5% din coeficienți în spațiul transformatei; (b) Transformata cosinus discretă a semnalului original: se observă că majoritatea coeficienților au valori neglijabile.

⁴S-a folosit rezultatul conform căruia, dacă $\varphi \in \mathbb{R}_{N \times 1}$ este vector propriu al unei matrici oarecare $\mathbf{A} \in \mathbb{R}_{N \times N}$, atunci el este vector propriu al oricărei matrici $\mathbf{A}^n$, cu $n \in \mathbb{Z}$. Într-adevăr, dacă avem $\mathbf{A}\varphi = \lambda\varphi$, atunci, înmulțind la stânga cu $\mathbf{A}$ avem: $\mathbf{A}^2\varphi = \mathbf{A}(\mathbf{A}\varphi) = \lambda\mathbf{A}\varphi = \lambda^2\varphi$ etc. Pentru puteri negative, înmulțim la stânga cu $\mathbf{A}^{-1}$ și obținem: $\mathbf{A}^{-1}\mathbf{A}\varphi = \lambda\mathbf{A}^{-1}\varphi$ ceea ce, prin împărțire la λ , conduce la $\mathbf{A}^{-1}\varphi = \lambda^{-1}\varphi$.

10.3.2 Transformata sinus discretă

Transformata sinus discretă este dată de matricea unitară $\mathbf{S} = \{s(k, n)\}_{k,n=0,\dots,N-1}$, având elementele date de:

$$s(k, n) = \sqrt{\frac{2}{N+1}} \sin \frac{\pi(k+1)(n+1)}{N+1}. \quad (10.49)$$

La rândul ei, și transformata sinus poate fi implementată prin algoritmi rapizi, de complexitate $\mathcal{O}(N \log N)$.

La fel ca și în cazul transformatei cosinus, se poate arăta că vectorii coloană ai matricii $\mathbf{S}^T$ sunt vectori proprii ai matricii:

$$\mathbf{N} = \begin{bmatrix} 1 & -\alpha & 0 & 0 & \cdots & 0 & 0 & 0 \\ -\alpha & 1 & -\alpha & 0 & \cdots & 0 & 0 & 0 \\ 0 & -\alpha & 1 & -\alpha & \cdots & 0 & 0 & 0 \\ 0 & 0 & -\alpha & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & -\alpha & 0 \\ 0 & 0 & 0 & 0 & \cdots & -\alpha & 1 & -\alpha \\ 0 & 0 & 0 & 0 & \cdots & 0 & -\alpha & 1 \end{bmatrix}. \quad (10.50)$$

Prin compararea (10.50) cu (10.44), se poate scrie:

$$\mathbf{N} \underset{|\rho| \leq 0,5}{\approx} \mathbf{K}_\xi^{-1}, \quad (10.51)$$

ceea ce se traduce prin faptul că transformata sinus discretă reprezintă o bună aproximare a transformatei optimale KL pentru semnale aleatoare slab corelate.

Capitolul 11

Cuantizarea semnalelor

11.1 Introducere

Aşa după cum am enunţat şi în capitolele anterioare, cuantizarea, împreună cu eşantionarea, reprezintă cei doi paşi prin care un semnal continuu este adus într-o formă în care poate fi prelucrat cu calculatorul. În urma operaţiei de eşantionare, semnalul continuu este transformat într-o secvenţă discretă de valori continue. Cuantizarea urmăreşte transformarea naturii valorilor eşantioanelor semnalului din continuu în discret. În acest scop, se foloseşte o funcţie de cuantizare Q cu care se aproximează valorile continue ale semnalului cu valori ce provin dintr-o mulţime cu L elemente (deci finită).

Considerând că gama de valori a semnalului este limitată în intervalul $[\xi_{\min}, \xi_{\max}]$ (ambele limite fiind presupuse cunoscute a priori) funcţia de cuantizare Q este de forma:

$$y = Q(x) = \begin{cases} y_L & \text{dacă } x_L < x \leq x_{L+1} \\ y_{L-1} & \text{dacă } x_{L-1} < x \leq x_L \\ \dots & \\ y_1 & \text{dacă } x_1 \leq x \leq x_2 \end{cases}, \quad (11.1)$$

unde $x_1 \equiv \xi_{\min}$ iar $x_{L+1} \equiv \xi_{\max}$. Valorile x_i se numesc *praguri de cuantizare*, în timp ce valorile y_i se numesc *valori cuantizate*. Diferenţa între două praguri de cuantizare succesive se notează cu $\Delta_i \stackrel{\text{not}}{=} x_i - x_{i-1}$ şi se numeşte *pas de cuantizare*. În figura 11.1 este prezentată forma generală a graficului unei funcţii de cuantizare.

Astfel, după cuantizare, fiecare eşantion al semnalului cuantizat

$$\dot{\xi}(k) = Q(\xi(k)) \quad (11.2)$$

poate fi codat pe un număr finit de biţi n_b cu $2^{n_b} \geq L$. În figura 11.2 sunt prezentate forma eşantioanelor unui semnal înainte şi după cuantizare.

Determinarea unei reguli de cuantizare constă, deci, în precizarea celor $2L - 1$ necunoscute, respectiv pragurile de cuantizare $x_2, \dots, x_{L-1}$ şi valorile cuantizate $y_1, \dots, y_L$. În paragrafele următoare vom prezenta câteva moduri de calcul al acestor valori.

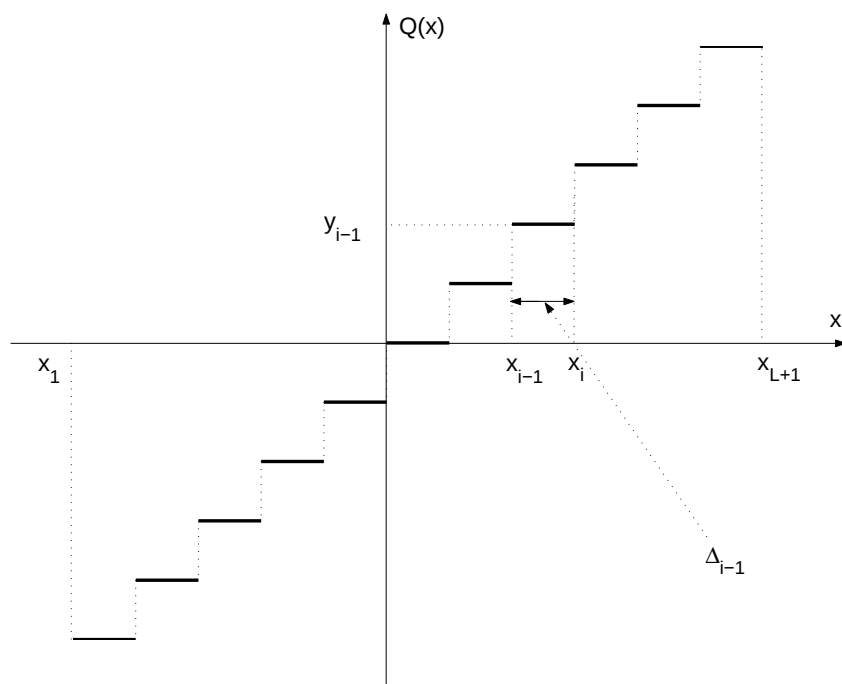


Figura 11.1: Graficul unei funcții de cuantizare $Q(x)$.

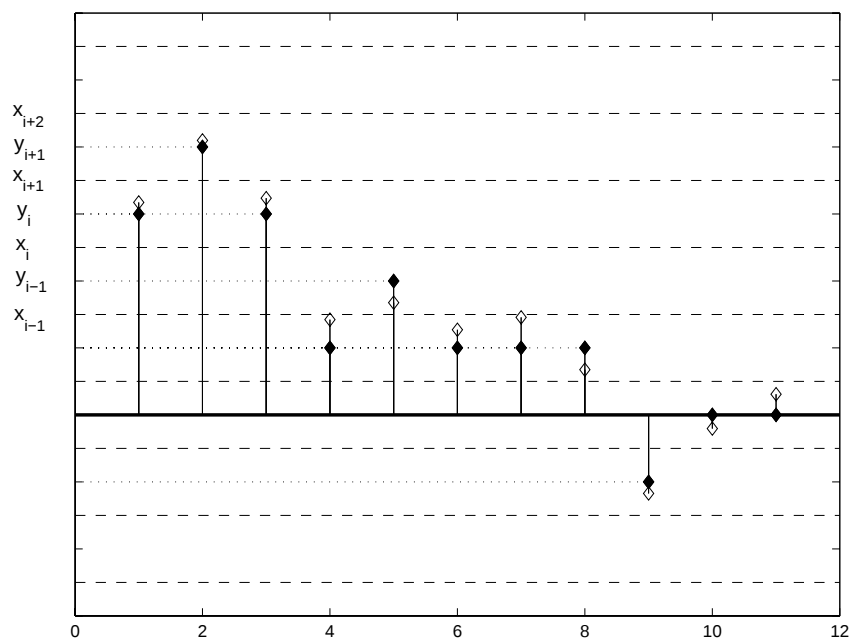


Figura 11.2: Cuantizarea valorilor unui semnal: cu $\diamond$ sunt figurate valorile originale $\xi(k)$ ale eșantioanelor semnalului, în timp ce valorile cuantizate $\hat{\xi}(k) = Q(\xi(k))$ sunt reprezentate cu $\blacklozenge$.

11.2 Cuantizarea uniformă

Cea mai simplă cuantizare este cea uniformă, care rezultă prin alegerea unui pas de cuantizare uniform pe tot intervalul de valori posibile ale semnalului $[\xi_{\min}, \xi_{\max}]$:

$$\Delta_{n-1} = \Delta_n \stackrel{\text{not}}{=} \Delta \quad \forall n = 2, 3, \dots, L. \quad (11.3)$$

Rezultă, deci, că

$$\Delta = \frac{\xi_{\max} - \xi_{\min}}{L}, \quad (11.4)$$

de unde rezultă valoarea pragurilor de cuantizare:

$$x_n = \xi_{\min} + (n-1)\Delta, \quad \forall n = 2, \dots, L. \quad (11.5)$$

Cât despre valorile cuantizate, ele se aleg la mijlocul fiecărui interval de cuantizare în parte:

$$y_n = \frac{x_n + x_{n+1}}{2} \quad \forall n = 1, 2, \dots, L. \quad (11.6)$$

Dacă se consideră eroarea indusă în semnal de operația de cuantizare:

$$e(n) = \xi(n) - \dot{\xi}(n), \quad (11.7)$$

atunci aceasta poate fi scrisă sub forma

$$e(n) = g(\xi(n)), \quad (11.8)$$

cu funcția $g(x)$ având forma:

$$g(x) = x - Q(x) = \begin{cases} x - y_L & \text{dacă } x_L < x \leq x_{L+1} \\ x - y_{L-1} & \text{dacă } x_{L-1} < x \leq x_L \\ \dots & \\ x - y_1 & \text{dacă } x_1 \leq x \leq x_2 \end{cases}. \quad (11.9)$$

Graficul funcției g corespunzătoare unei cuantizări uniforme este prezentat în figura 11.3.

Aplicând (3.25), rezultă

$$w_e(u) = \begin{cases} \sum_{i=1}^L w_{\xi}(u + y_i) & \text{dacă } u \in \left[-\frac{\Delta}{2}, \frac{\Delta}{2}\right] \\ 0 & \text{în rest} \end{cases}. \quad (11.10)$$

Se poate arăta că dacă numărul de nivele de cuantizare L este suficient de mare pentru o gamă de valori precizată $[\xi_{\min}, \xi_{\max}]$, ceea ce este echivalent cu a spune că pasul de cuantizare (11.4) este suficient de mic, distribuția erorii induse prin cuantizare uniformă este aproximativ uniformă în intervalul precizat:

$$w_e(u) \underset{\Delta \searrow}{\approx} \begin{cases} \frac{1}{\Delta} & \text{dacă } |u| \leq \frac{\Delta}{2} \\ 0 & \text{în rest} \end{cases}. \quad (11.11)$$

Rezultă din (11.11) că eroarea pătratică medie indusă de cuantizarea uniformă, care nu este altceva decât varianța erorii $e(n)$ este dată de:

$$\varepsilon = \overline{e^2(n)} = \sigma_e^2 = \frac{\Delta^2}{12}. \quad (11.12)$$

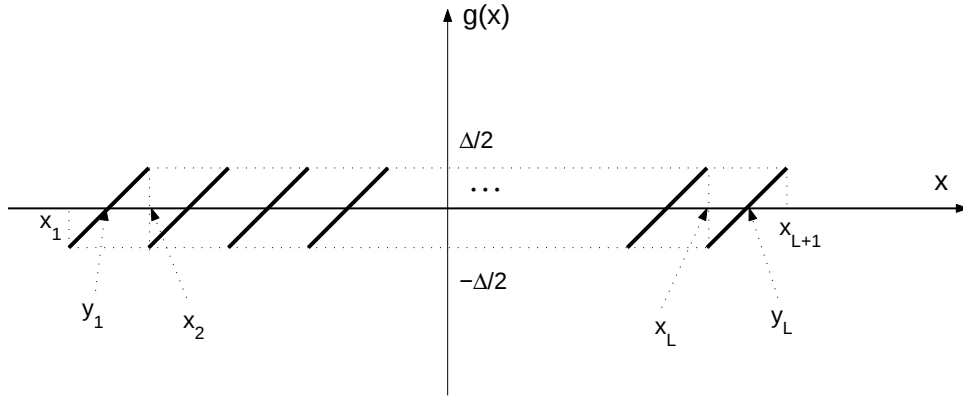


Figura 11.3: Funcția de trecere din ξ în e

11.3 Cuantizarea optimală Lloyd–Max

Cuantizarea uniformă dată de relațiile (11.5) și (11.6) este atractivă prin prisma simplității modului de calcul al pragurilor de cuantizare. Totuși ea nu este optimală, întrucât nu ține cont de distribuția valorilor semnalului de cuantizat w_ξ . Cuantizarea Lloyd–Max urmărește calculul pragurilor de cuantizare și al valorilor cuantizate în sensul minimizării erorii pătratice medii induse de cuantizare. În mare, ideea este de a alocă pași de cuantizare Δ_i mici pe intervalele pe care distribuția w_ξ are valori importante, și, implicit, de a alocă pași de cuantizare mai mari (respectiv erori mai mari) pentru intervalele pe care w_ξ este mai mică (respectiv pentru valorile mai puțin probabile).

Eroarea pătratică medie, se scrie ca:

$$\varepsilon = \overline{e^2(n)} = \overline{g(\xi(n))^2} = \int_{-\infty}^{\infty} g^2(x) w_\xi(x) dx. \quad (11.13)$$

În dezvoltarea de mai sus am aplicat teorema de medie (3.41) și am ținut cont că semnalul $\xi(n)$ este presupus din oficiu a fi staționar, deci este caracterizat de aceeași densitate de probabilitate w_ξ la orice moment de timp n . Înlocuind pe $g(x)$ cu relația dată de (11.9)¹, avem

$$\varepsilon = \sum_{i=1}^L \int_{x_i}^{x_{i+1}} (x - y_i)^2 w_\xi(x) dx. \quad (11.14)$$

Valorile pragurilor x_k ce minimizează pe ε se obțin din ecuația:

$$\frac{\partial \varepsilon}{\partial x_k} = 0 \quad \forall k = 2, 3, \dots, L, \quad (11.15)$$

ecuație care, ținând cont de (4.9), se scrie:

$$2(x_k - y_{k-1})^2 w_\xi(x_k) - 2(x_k - y_k)^2 w_\xi(x_k) = 0. \quad (11.16)$$

¹Formula este valabilă pentru orice lege de cuantizare, adică pentru orice valori ale lui x_i și y_i .

După reduceri și rearanjarea termenilor, obținem:

$$x_k = \frac{y_{k-1} + y_k}{2} \quad \forall k = 2, \dots, L. \quad (11.17)$$

Pe de altă parte, valorile cuantizate y_k ce minimizează eroarea pătratică medie ε se obțin, la rândul lor, din ecuația:

$$\frac{\partial \varepsilon}{\partial y_k} = 0 \quad \forall k = 2, 3, \dots, L, \quad (11.18)$$

care este echivalentă cu:

$$-2 \int_{x_k}^{x_{k+1}} (x - y_k) w_\xi(x) dx = 0. \quad (11.19)$$

Prin rearanjarea termenilor, obținem valorile finale ale lui y_k :

$$y_k = \frac{\int_{x_k}^{x_{k+1}} x w_\xi(x) dx}{\int_{x_k}^{x_{k+1}} w_\xi(x) dx} \quad \forall k = 1, 2, \dots, L. \quad (11.20)$$

Ținând cont de relația (3.10), putem scrie expresia valorilor cuantizate optime sub forma:

$$y_k = \int_{-\infty}^{\infty} x w_{\xi|\{x_k \leq \xi \leq x_{k+1}\}}(x) dx = \bar{\xi}|\{x_k \leq \xi \leq x_{k+1}\} \quad (11.21)$$

Relațiile (11.17) și (11.20) formează sistemul de ecuații Lloyd–Max, care este un sistem neliniar de $2L-1$ ecuații cu $2L-1$ necunoscute prin a cărui rezolvare se determină pragurile de cuantizare și valorile cuantizate ce minimizează eroarea pătratică medie de cuantizare.

Să mai spunem că, în general, nu există o soluție analitică a sistemului Lloyd–Max, și că aceasta se află, în general, prin metode numerice. Totuși, după cum vom arăta în paragraful 11.4, pentru cazul particular al unui număr de nivele de cuantizare L ridicat, soluția sistemului Lloyd–Max poate fi dată în mod analitic cu bună aproximație.

Pentru moment, să ne oprim asupra observației că, în cazul în care semnalul de cuantizat are o distribuție uniformă în intervalul $[\xi_{\min}, \xi_{\max}]$, atunci, ținând cont de (3.22), ecuația (11.20) se scrie:

$$y_k = \frac{\int_{x_k}^{x_{k+1}} x \frac{1}{\xi_{\max} - \xi_{\min}} dx}{\int_{x_k}^{x_{k+1}} \frac{1}{\xi_{\max} - \xi_{\min}} dx} = \frac{\frac{x^2}{2} \Big|_{x_k}^{x_{k+1}}}{x \Big|_{x_k}^{x_{k+1}}} = \frac{x_k + x_{k+1}}{2}. \quad (11.22)$$

În continuare, (11.17) devine:

$$x_k = \frac{\frac{x_{k-1} + x_k}{2} + \frac{x_k + x_{k+1}}{2}}{2}, \quad (11.23)$$

care, prin operații elementare, poate fi scrisă ca:

$$x_{k+1} - x_k = x_k - x_{k-1} \quad \forall k = 2, \dots, L. \quad (11.24)$$

Comparând (11.23) cu (11.3), rezultă că, în cazul unei distribuții uniforme a semnalului de cuantizat, cuantizorul optimal este chiar cel uniform.

11.4 Compandarea

Problema pe care ne-o punem este de a implementa cuantizarea optimală folosind un dispozitiv fizic de cuantizare uniformă. Trebuie, deci, găsită o funcție $f(x)$, astfel încât prin succesiunea de operații

- transformare:

$$\eta(n) = f(\xi(n)) \quad (11.25)$$

- cuantizare uniformă:

$$\dot{\eta}(n) = Q(\eta(n)) \quad (11.26)$$

cu Q dat de (11.5) și (11.6),

- transformare inversă:

$$\dot{\xi}(n) = f^{-1}(\dot{\eta}(n)) \quad (11.27)$$

să se obțină o cuantizare optimală a semnalului $\xi(n)$. Datorită faptului că alura funcției f rezultate pentru majoritatea distribuțiilor întâlnite în practică este aceea a unei funcții de compresie a gamei semnalului (vezi figura 11.5), și deci, implicit, funcția f^{-1} realizează o expandare, tehnica poartă numele de *compandare*, termen ce provine din prescurtarea sintagmei *compresie-expandare*. Schema bloc a cuantizării prin compandare este prezentată în figura (11.4).

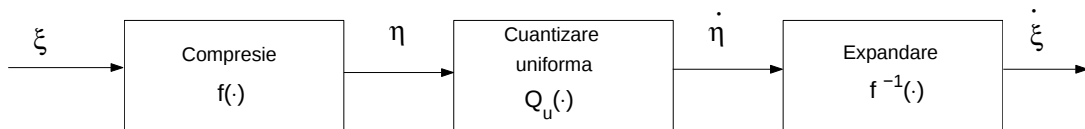


Figura 11.4: Cuantizarea prin “compandare”.

Cât despre determinarea funcției $f(x)$, ne vom mărgini la a da expresia sa aproximativă, obținută în ipoteza unui număr de nivele L suficient de mare ca densitatea de probabilitate w_ξ să poată fi considerată aproximativ constantă pe orice interval de cuantizare:

$$w_\xi(x) \underset{L \gg}{\approx} \text{const} = w_\xi(y_k) \quad \forall x \in [x_k, x_{k+1}], \quad \forall k = 1, \dots, L. \quad (11.28)$$

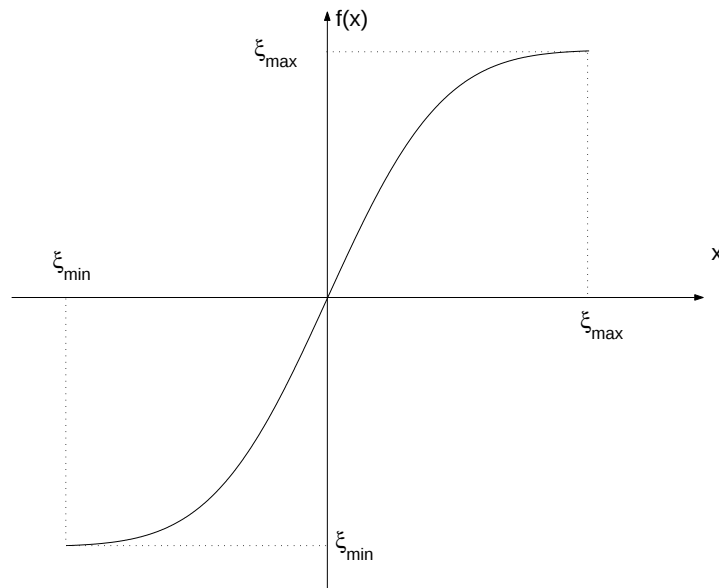


Figura 11.5: Forma funcției de “compresie” din relația (11.29) pentru o distribuție w_ξ de tip Gauss.

În această ipoteză, funcția $f(x)$ poate fi aproximată cu:

$$f(x) = (\xi_{\max} - \xi_{\min}) \frac{\int_{-\xi_{\min}}^x [w_\xi(u)]^{\frac{1}{3}} du}{\int_{-\xi_{\min}}^{\xi_{\max}} [w_\xi(u)]^{\frac{1}{3}} du} + \xi_{\min}. \quad (11.29)$$

Calculul ce permite deducerea relației (11.29) este prezentat în referința [9]. În figura 11.5 este prezentată forma funcției $f(x)$ pentru o distribuție w_ξ gaussiană. Se poate observa forma funcției, având o pantă ce scade spre extreme, ceea ce justifică denumirea de funcție de “compresie”.

Capitolul 12

Aplicații în prelucrarea și analiza imaginilor

12.1 Introducere

Unul dintre domeniile cu aplicații dintre cele mai spectaculoase ale teoriei și modelelor statistice prezentate în capitolele anterioare este prelucrarea și analiza digitală a imaginilor. Imaginile, în accepțiunea cea mai simplă, sunt semnale bidimensionale – funcții dependente de două variabile a căror semnificație este de coordonate spațiale. Valorile unei imagini provin din achiziția din scena reală a unei mărimi fizice de interes de către un senzor adaptat corespunzător. Reprezentarea digitală a imaginilor înseamnă reprezentarea valorilor acestora într-o formă utilizabilă direct de către un calculator digital, deci o reprezentare cu un număr finit de biți, ce rezultă în urma eșantionării și cuantizării¹ semnalului continuu preluat de senzor.

Deci, pentru o imagine digitală, valorile posibile sunt în număr finit (în general corespunzând reprezentării valorilor cuantizate ca întregi, printr-un număr fixat de biți), iar variabilele ce corespund suportului spațial (coordonatele în spațiul imaginii, sau variabilele de care depinde funcția imagine) au la rândul lor valori discrete.

Putem deci considera în cazul cel mai simplu că o imagine digitală este o matrice (sau tablou) în care sunt înscrise valorile funcției imagine. Fiecare element al acestei matrici, deci o celulă situată la intersecția unei linii și a unei coloane date, se numește pixel (termenul folosit în limba română fiind identic cu cel folosit în limba engleză, unde provine din contragerea termenilor *picture element* – element de imagine). Așadar, o imagine digitală este alcătuită din pixeli, fiecare pixel fiind caracterizat de o anumită poziție (coordonate de linie și de coloană corespunzătoare plasamentului său în matricea imagine) și de o anumită valoare².

¹De cele mai multe ori cuantizarea este uniformă, numărul suficient de mare de nivele de cuantizare asigurând precizia necesară (eroarea suficient de mică) de reprezentare.

²Interesează de asemenea forma pixelului (evident pătrată în exemplificarea de mai sus, dar putând avea și alte forme – de exemplu hexagonală – în aplicații mai complexe).

În covârșitoarea majoritate a aplicațiilor, imaginile sunt de natură vizuală, corespunzând măsurării în scena reală a unei intensități medii a radiației electromagnetice din domeniul spectral vizibil – luminanța. Imaginea rezultantă corespunde unei imagini fotografice alb-negru (sau unei imagini de televiziune alb-negru); valorile pixelilor codează (reprezintă) luminozitatea punctelor corespunzătoare din scenă, între o valoare minimă nulă, ce corespunde tonului celui mai întunecat (negru) și o valoare maximă $M - 1$, determinată de numărul de biți folosiți pentru reprezentarea binară a valorilor ($M = 2^B$), ce corespunde tonului celui mai deschis sau strălucitor (alb). Imaginea se numește cu nivele (sau tonuri) de gri. S-a demonstrat că pentru aplicații de tipul transmisiei de televiziune sunt suficiente $M = 256$ de nivele de gri (și deci o reprezentare digitală pe $B = 8$ biți) pentru a asigura o percepție vizuală de calitate suficientă.



Figura 12.1: Imagini de luminanță, cuantizate cu 256 nivele de gri: imagine naturală și fragment dintr-o radiografie de calcaneu.

Odată ce imaginea a fost achiziționată din scenă și a fost eșantionată și cuantizată (deci reprezentată în formă digitală) mulțimea de valori ale pixelilor poate fi stocată și prelucrată de un sistem digital de calcul. Succesiunea de operații prin care se poate modifica aspectul vizual sau forma de codare a imaginii sau prin care se pot extrage caracteristici de interes necesare luării unor decizii formează obiectul prelucrării și analizei imaginilor.

Ca și în cazul semnalelor unidimensionale, o bună parte din tehnicile de prelucrare și analiză a imaginilor sunt bazate pe abordarea statistică, abordare ce presupune considerarea unei imagini ca fiind o *realizare particulară* a unui semnal aleator cu suport bidimensional (numit câmp aleator) care este caracterizat de mărimi statistice. În multe cazuri, din cauza faptului că nu dispunem de suficient de multe imagini aparținând aceleiași clase, trebuie să facem ipoteza de ergodicitate a câmpului aleator, ipoteză ce permite determinarea mărimilor respective pe baza valorilor pixelilor dintr-o singură imagine.

Mărimea statistică cel mai des folosită este densitatea de probabilitate de ordinul unu a câmpului aleator. Datorită discretizării valorilor pixelilor din imagine, după cum am arătat în paragraful 3.4, densitatea de probabilitate căutată este nulă peste tot cu excepția punctelor ce corespund valorilor cuantizate ale pixelilor, $\{0, 1, \dots, M - 1\}$. În

aceste puncte, funcția de densitate de probabilitate acumulează probabilitățile de apariție ale valorilor corespunzătoare, probabilități ce trebuie estimate conform (2.1). În ipoteza că valorile pixelilor dintr-o imagine sunt independente statistic, atunci respectivele probabilități pot fi estimate folosind valorile pixelilor *din imagine*. Funcția rezultată, care are semnificația de densitate de probabilitate de ordinul unu determinată experimental, se numește *histograma imaginii*, și este dată de:

$$h(i) = \frac{N_i}{L \times C} \quad i = 0, 1, \dots, M-1, \quad (12.1)$$

unde N_i este numărul de pixeli din imagine ce au nivelul de gri i , iar L și C reprezintă dimensiunile imaginii (respectiv numărul de linii și de coloane ale matricii imagine).

Evident, după modul în care a fost definită, histograma oricărei imagini verifică condiția de normare:

$$\sum_{i=0}^{M-1} h(i) = 1. \quad (12.2)$$

ceea ce este normal, fiind vorba de o funcție asimilabilă unei densități de probabilitate.

În figura 12.2 sunt prezentate histogrammele imaginilor din figura 12.1.

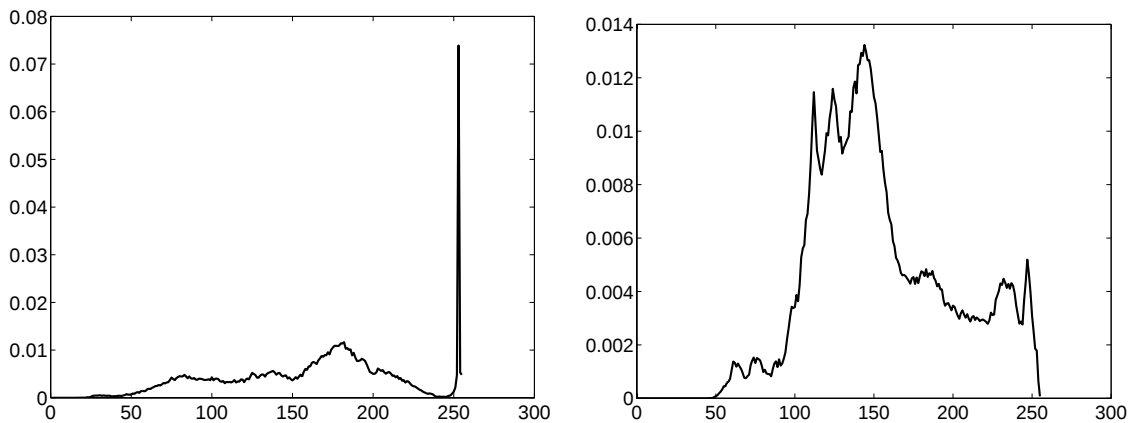


Figura 12.2: Histogrammele imaginilor din figura 12.1. Abscisele i ale graficelor reprezintă nivelul de gri, iar ordonatele $h(i)$ sunt probabilitățile de apariție a nivelelor de gri i . Graficele au fost reprezentate presupunând variația continuă a nivelelor de gri.

Histograma imaginii oferă informații asupra conținutului acesteia: gama de variație a valorilor, calitatea perceptuală a imaginilor, numărul de tipuri de obiecte conținute în scenă. Fiind însă o distribuție de ordinul unu, histograma nu descrie așezarea spațială a valorilor pixelilor, fiind astfel posibil ca imagini cu conținut vizual extrem de diferit să fie descrise de histogramme asemănătoare.

În cele ce urmează vom prezenta câteva metode ce utilizează histograma imaginii (deci densitatea de probabilitate de ordinul unu a valorilor pixelilor) pentru realizarea îmbunătățirii imaginii și a segmentării (binarizării) imaginii.

12.2 Îmbunătățirea imaginilor

Îmbunătățirea imaginilor este o sintagmă generală ce se referă la o clasă largă de operații al căror scop este mărirea vizibilității componentelor (sau părților) imaginii, în sensul percepției vizuale umane. Ca urmare, criteriile de evaluare ale calității unei imagini vor fi subiective (depinzând de utilizatorul imaginii) și specifice aplicației din care provine imaginea. Se poate face analogia operațiilor de îmbunătățire a imaginilor cu reglajul tonalității muzicii ascultate; astfel, în funcție de ascultător, se pot favoriza componentele înalte sau joase, sau nici unele. În principiu îmbunătățirea imaginilor se poate face fără a lua în considerare nici o informație asupra imaginii originale sau asupra procesului de degradare care a determinat ca imaginea dată să necesite o îmbunătățire. Conform acestui punct de vedere chiar și o imagine originală (nedegradată) poate fi îmbunătățită, obținând o imagine falsificată (în sensul diferențelor față de original), dar subiectiv preferabilă (un contrast mai mare, muchii și frontiere mai accentuate, regiuni uniforme mai netede).

Prin îmbunătățire, unei imagini nu i se adaugă nici o informație nouă față de cea existentă inițial; conținutul original este însă prezentat diferit. Deși la o examinare superficială afirmația este corectă, putem găsi măcar două obiecții (sau contraexemple) la această formulare:

- din punctul de vedere al utilizatorului, informația, chiar dacă există, nu poate fi folosită, deci este asimilabil nulă. Acesta este cazul imaginilor obținute în condiții extreme de iluminare, ce prezintă un contrast foarte slab (imagini subexpuse sau supraexpuse);
- din punctul de vedere al teoriei informației, informația din imagine poate fi asimilată entropiei procesului aleator ale cărui realizări particulare sunt valorile de gri ale pixelilor; entropia se modifică însă la orice transformare ce afectează distribuția nivelelor de gri din imagine.

12.2.1 Egalizarea de histogramă

Majoritatea imaginilor prezintă o distribuție neuniformă a nivelelor de gri – în orice imagine există nivele de gri predominante și există nivele de gri puțin prezente sau absente. Operațiile de îmbunătățire a imaginilor (pentru îmbunătățirea percepției vizuale) au ca scop redistribuirea nivelelor de gri din imagine, astfel încât acestea să ocupe întreaga gamă de valori disponibile $[0, M - 1]$. Dacă se impune condiția ca noua distribuție a nivelelor de gri să fie uniformă, operația ce realizează această cerință se numește egalizare de histogramă.

Scopul egalizării de histogramă este deci obținerea unei distribuții uniforme a nivelelor de gri; imaginea rezultată va prezenta cea mai mare îmbunătățire a contrastului, contrast distribuit regulat în întreaga gamă de valori posibile a nivelelor de gri. Din punct de vedere matematic, egalizarea de histogramă înseamnă transformarea unei distribuții oarecari (descrisă de histograma imaginii inițiale) într-o distribuție uniformă. Implementarea

egalizării de histogramă presupune deci determinarea acestei funcții scalare de o variabilă (care modifică valorile nivelelor de gri).

După cum s-a arătat în exemplul de la pagina 27, pentru orice variabilă aleatoare (având deci orice distribuție) funcția sa de repartiție o transformă într-o variabilă aleatoare distribuită uniform în intervalul $[0, 1]$. În ceea ce privește funcția de repartiție de ordinul unu al imaginii, aceasta se calculează, conform proprietății 2 a densităților de probabilitate, după relația:

$$H(i) = \sum_{j=0}^i h(j) \quad i = 0, 1, \dots, M-1. \quad (12.3)$$

și se numește *histogramă cumulativă*.

Astfel, egalizarea de histogramă se obține modificând valoarea fiecărui pixel din imagine printr-o funcție ce se obține prin rescalarea valorilor histogramei cumulative a imaginii (12.3) din intervalul $[0, 1]$ în toată gama de nivele de gri ale imaginii, $[0, M-1]$. Rezumând, putem scrie expresia funcției de realocare a nivelelor de gri $F(i)$ cu care se implementează operația de egalizare de histogramă ca:

$$F(i) = \left\lfloor \frac{H(i) - H(0)}{1 - H(0)} (M-1) \right\rfloor, \quad i = 0, 1, \dots, M-1. \quad (12.4)$$

unde cu $\lfloor \cdot \rfloor$ am notat operatorul de rotunjire la cel mai apropiat întreg. Figurile 12.3 prezintă efectul egalizării de histogramă asupra unor imagini naturale.

Egalizarea de histogramă este deci o metodă automată (nu necesită stabilirea de parametri) și adaptivă de îmbunătățire a aspectului vizual al imaginilor. Discuția asupra limitărilor și deficiențelor acestei metode depășește însă cadrul acestui material – cititorul interesat fiind invitat a se referi la un curs dedicat prelucrării și analizei imaginilor, precum [6].

12.3 Segmentarea imaginilor

Operația de segmentare³ urmărește extragerea din imagine a zonelor (regiunilor) ocupate de diferitele obiecte prezente în scenă. Un obiect se definește ca o entitate caracterizată de un set de parametri ale căror valori nu se modifică în diferitele puncte ce aparțin entității considerate. Mai simplu, putem spune că obiectul are proprietatea de uniformitate a parametrilor de definiție.

Unul dintre cei mai simpli parametri de definiție este chiar nivelul de gri. După cum am menționat, nivelul de gri corespunde în scenă unei proprietăți fizice ce este preluată de senzorul de imagine. În acest caz, histograma imaginii reflectă distribuția în scenă a proprietății fizice înregistrate. Dacă nivelul de gri (respectiv proprietatea fizică pe care acesta o reprezintă) caracterizează în mod suficient obiectele din scenă, histograma imaginii va prezenta o structură de moduri dominante - intervale disjuncte de nivele de

³În acest material ne vom referi numai la metode de segmentare zise “orientate pe regiuni” și nu la extragerea conturilor obiectelor din imagini.



Figura 12.3: Coloana din stânga: imagini cu nivele de gri; coloana din dreapta: imagini după egalizarea de histogramă.

gri ce apar cu probabilitate mai mare. Fiecare asemenea mod (maxim al histogramei) va reprezenta o anumită categorie de obiecte și vom numi “clasă” intervalul de nivele de gri alocat unui anume tip de obiect.

Ca exemple imediate putem menționa imaginile obținute prin scanarea documentelor scrise și a tipăriturilor (folosite de programe de tip OCR - *Optical Character Recognition*), imaginile în infraroșu (nivelul de gri al unui pixel este asociat temperaturii punctului respectiv, astfel încât “mai alb” are semnificația de “mai cald”) sau imaginile în care există o categorie de obiecte de interes și un fundal uniform. Pentru toate aceste tipuri de imagini histograma este de tipul celei prezentate în figura 12.4. În general, imaginile reale (zise naturale) nu prezintă o asemenea structură clară de moduri (după cum arată și exemplele din figura 12.2).

Segmentarea pe histogramă (numită și prăguire sau *thresholding*) urmărește deter-

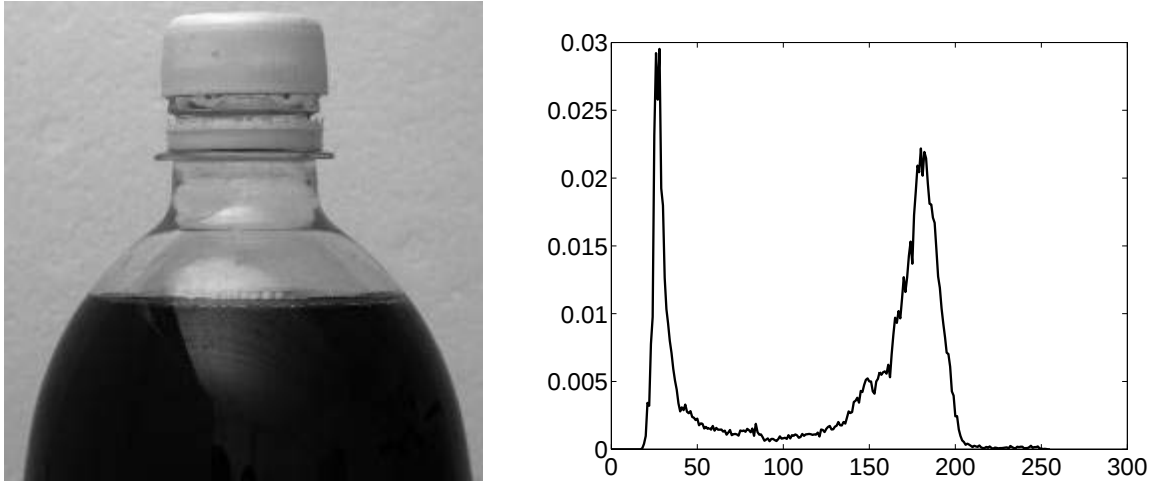


Figura 12.4: Exemplu de imagine cu nivele de gri având o histogramă bimodală, cu moduri bine separate.

minarea unor nivele de gri ce separă modurile histogramei. Apoi, tuturor punctelor din imagine al căror nivel de gri corespunde aceluiași mod, li se asociază o etichetă comună, a cărei semnificație este de apartenență a punctului la obiectul respectiv. Pentru exemplele prezentate anterior, problema este una de segmentare în două clase (respectiv împărțirea pixelilor imaginii în caractere scrise/zone albe, obiecte fierbinți/obiecte reci, obiect/fundal) operație ce necesită determinarea unui singur nivel de gri de separație T , numit *prag de segmentare*. Acest prag de segmentare poate fi ales pe minimul histogramei ce separă modurile acesteia (dacă separația este evidentă). Astfel, pornind de la imaginea inițială f , imaginea etichetată g va fi dată de:

$$g(m, n) = \begin{cases} E_0 & \text{dacă } 0 \leq f(m, n) < T \\ E_1 & \text{dacă } T \leq f(m, n) < M \end{cases} \quad (12.5)$$

Imaginea etichetată g – rezultatul transformării – va fi descrisă de cele două etichete: eticheta E_0 pentru punctele al căror nivel de gri este mai mic decât pragul T și eticheta E_1 pentru punctele al căror nivel de gri este mai mare decât pragul T . Etichetele E_0 și E_1 pot fi valori numerice (0 și 1, sau 0 și 255) sau pot fi șiruri de simboluri. Acest tip de segmentare se numește binarizare, întrucât imaginea rezultat este caracterizată de numai două valori.

În cazul general al existenței mai multor obiecte în imagine, trebuie ales un număr corespunzător de praguri de segmentare T_k , segmentarea fiind descrisă de:

$$g(m, n) = E_k \quad \text{dacă } T_k \leq f(m, n) < T_{k+1} \quad (12.6)$$

unde $T_0 = 0$, $T_C = M$ și $k = 0, 1, \dots, C - 1$. Pragurile T_k pot fi alese manual, prin inspecția histogramei, în minimele locale ale acesteia.

Una din cerințele de bază în majoritatea aplicațiilor din prelucrările de imagine este ca acestea să lucreze în mod automat, fără intervenția din exterior a utilizatorului (care



Figura 12.5: Imagini segmentate (binarizate): stânga – varianta binarizată cu un prag $T = 110$ a imaginii cu nivele de gri din figura 12.4; dreapta – varianta binarizată cu un prag $T = 190$ a imaginii cu nivele de gri din figura 12.3.d. Eticheta E_0 este reprezentată cu negru, iar eticheta E_1 este reprezentată cu alb.

nu este, în general, de specialitate). Se pune, deci, problema alegerii automate a pragului(urilor) de segmentare T_k , pe baza analizei automate a histogramei imaginii și a eventualelor cunoștințe a priori asupra conținutului imaginilor. În paragrafele următoare, vom prezenta două abordări posibile pentru binarizarea automată a unei imagini.

12.3.1 Segmentarea imaginilor ca problemă de cuantizare

După cum am arătat în capitolul 11, cuantizarea unor valori implică înlocuirea valorilor respective cu cea mai apropiată valoare dintr-o mulțime finită de valori (mulțimea valorilor cuantizate). Operația de cuantizare este o aproximare a valorilor inițiale prin valorile cuantizate, mai exact, aproximarea tuturor valorilor cuprinse în intervalul $(x_i, x_{i+1}]$ cu valoarea cuantizată y_i . În cuantizarea optimală Lloyd-Max (prezentată în paragraful 11.3) valorile cuantizate și intervalele de cuantizare sunt determinate astfel încât media pătratică a erorii induse de aproximarea valorilor inițiale cu cele cuantizate să fie minimă pentru un număr de nivele de cuantizare dat.

După cum s-a demonstrat (v. relația (11.21)), valorile cuantizate care minimizează eroarea pătratică medie de aproximare sunt mediile statistice ale valorilor din intervalul de cuantizare respectiv. Altfel spus, cuantizarea optimală înseamnă alegerea intervalelor de cuantizare sunt astfel încât valorile din fiecare interval să fie cât mai apropiate de media lor, adică varianța valorilor din fiecare interval de cuantizare să fie minimă. Eroarea de cuantizare va fi atunci varianța medie corespunzătoare tuturor intervalelor de cuantizare alese. Într-adevăr, înlocuind (11.21) în (11.14) putem scrie eroarea pătratică medie ca:

$$\begin{aligned}
\varepsilon &= \sum_{i=1}^L \int_{x_i}^{x_{i+1}} (x - \bar{\xi}|\{x_i \leq \xi \leq x_{i+1}\})^2 w_\xi(x) dx = \\
&= \sum_{i=1}^L \underbrace{\int_{x_i}^{x_{i+1}} w_\xi(z) dz}_{P(x_i \leq \xi \leq x_{i+1})} \underbrace{\int_{x_i}^{x_{i+1}} (x - \bar{\xi}|\{x_i \leq \xi \leq x_{i+1}\})^2 \frac{w_\xi(x)}{\int_{x_i}^{x_{i+1}} w_\xi(z) dz} dx}_{\sigma_{\xi|\{x_i \leq \xi \leq x_{i+1}\}}^2} \\
&= \sum_{i=1}^L P(x_i \leq \xi \leq x_{i+1}) \sigma_{\xi|\{x_i \leq \xi \leq x_{i+1}\}}^2
\end{aligned} \tag{12.7}$$

În calculul de mai sus am folosit rezultatul (3.10).

În cazul binarizării, problema (care fusese inițial formulată doar ca alegere a pragului de segmentare T care separă modurile histogramei ce corespund celor două obiecte din imagine) poate fi reformulată în sensul cuantizării astfel: dorim cuantizarea optimă a valorilor pixelilor imaginii cu $L = 2$ nivele, ceea ce revine la determinarea pragului T astfel încât prin împărțirea gamei de nivele de gri disponibile $[0, M - 1]$ în două subintervale, $[0, T]$ și $[T, M - 1]$, și prin aproximarea fiecărui pixel dintr-un subinterval cu media acestuia, eroarea pătratică medie să fie minimă. Astfel, conform (12.7), pragul T cu care facem binarizarea trebuie să minimizeze expresia:

$$\varepsilon = P_0 \sigma_0^2 + P_1 \sigma_1^2, \tag{12.8}$$

unde P_0 și P_1 sunt probabilitățile ca nivelul de gri al unui pixel să aparțină intervalelor $[0, T - 1]$, respectiv $[T, M - 1]$:

$$P_0 = \sum_{i=0}^{T-1} h(i), \tag{12.9a}$$

$$P_1 = \sum_{i=T}^{M-1} h_i = 1 - P_0, \tag{12.9b}$$

iar σ_0^2 și σ_1^2 sunt varianțele nivelelor de gri corespunzătoare aceluiași intervale:

$$\sigma_0^2 = \frac{\sum_{i=0}^{T-1} i^2 h(i)}{\sum_{i=0}^{T-1} h(i)} - \left(\frac{\sum_{i=0}^{T-1} i h(i)}{\sum_{i=0}^{T-1} h(i)} \right)^2, \tag{12.10a}$$

$$\sigma_1^2 = \frac{\sum_{i=T}^{M-1} i^2 h(i)}{\sum_{i=T}^{M-1} h(i)} - \left(\frac{\sum_{i=T}^{M-1} i h(i)}{\sum_{i=T}^{M-1} h(i)} \right)^2. \tag{12.10b}$$

Acest mod de definire a binarizării este cunoscut în domeniul prelucrării și analizei imaginilor drept metoda de segmentare Otsu.

După cum am arătat, pragul căutat T este media aritmetică a mediilor statistice ale valorilor de cuantizat din cele două intervale de cuantizare alese; ținând seama că nivelele de gri din imagine sunt (deja) valori discrete, expresia (11.17) se poate scrie ca:

$$T = \frac{1}{2} \left(\frac{\sum_{i=0}^{T-1} ih(i)}{\sum_{i=0}^{T-1} h(i)} + \frac{\sum_{i=T}^{M-1} ih(i)}{\sum_{i=T}^{M-1} h(i)} \right). \quad (12.11)$$

$\mu_0 \qquad \mu_1$

Această ecuație nu se poate rezolva analitic. Putem găsi totuși valoarea T căutată, care minimizează pe (12.8) calculând valorile unei forme modificate ale funcției (12.8) pentru toate valorile posibile ale lui $T \in \{0, 1, \dots, M-1\}$. Pentru aceasta, 12.8 se poate rescrie ca:

$$\varepsilon = \sum_{i=0}^{M-1} i^2 h(i) - (P_0 \mu_0^2 + P_1 \mu_1^2). \quad (12.12)$$

Minimizarea lui (12.8) revine la maximizarea lui $P_0 \mu_0^2 + P_1 \mu_1^2$, cu μ_0 și μ_1 mediile nivelelor de gri pe cele două intervale. O simplificare a calculului se poate obține prin considerarea relațiilor iterative prin care mediile și probabilitățile de apariție pe intervale, calculate pentru pragul $T+1$ depind de aceleași mărimi calculate pentru pragul imediat anterior, T . De exemplu avem:

$$P_0(T+1) = P_0(T) + h(T+1) \quad (12.13)$$

$$\mu_0(T+1) = \frac{P_0(T)\mu_0(T) + (T+1)h(T+1)}{P_0(T+1)}. \quad (12.14)$$

Relații analoage există pentru mărimile P_1 și μ_1 .

12.3.2 Segmentarea imaginilor ca problemă de decizie

Binarizarea unei imagini cu nivele de gri poate fi interpretată și ca o problemă de decizie – trebuie hotărâtă clasa (tipul de obiect) căreia îi aparține fiecare punct din imagine. Decizia se reduce, desigur, la calculul pragului de segmentare T , cu care se compară nivelele de gri ale pixelilor din imagine, astfel încât clasificarea pixelilor în clase să fie cât mai corectă din punctul de vedere al unui criteriu de eroare. Pragul astfel determinat va fi optim în sensul criteriului de eroare folosit, segmentarea numindu-se segmentare cu prag optim.

În cazul binarizării trebuie determinat deci un unic prag T ce separă modurile din histogramă ce corespund celor două obiecte din imagine: obiecte O_0 , caracterizate de nivele de gri mai mici decât pragul T , și obiecte O_1 , ale căror nivele de gri sunt mai mari

decât pragul T . Binarizarea se poate formula într-un cadru specific proceselor de decizie, considerând că pentru fiecare pixel al imaginii se ia decizia D_0 (respectiv se decide că pixelul aparține unui obiect O_0) dacă nivelul său de gri este mai mic decât pragul T , sau decizia D_1 (pixelul aparține unui obiect de tip O_1) dacă nivelul de gri al pixelului este mai mare decât pragul T . După cum se poate remarca, observația pe baza căreia luăm decizia este un scalar, și anume nivelul de gri al pixelului respectiv.

Criteriul ce se urmărește optimizat este minimizarea probabilității de clasificare eronată a punctelor din imagine. Presupunând cunoscute funcțiile de densitate de probabilitate a nivelelor de gri ce descriu fiecare dintre cele două tipuri de componente din imagine, $w_0(i)$ și $w_1(i)$ și probabilitățile lor de apariție în imagine, P_0 și P_1 , putem exprima probabilitatea de clasificare eronată ca:

$$P_E(T) = P_0 \underbrace{\int_T^{+\infty} w_0(x) dx}_{P(D_1|O_0)} + P_1 \underbrace{\int_{-\infty}^T w_1(x) dx}_{P(D_0|O_1)} \quad (12.15)$$

În mod evident se presupune că cele două tipuri de obiecte ocupă complet imaginea ($P_0 + P_1 = 1$). Histograma imaginii (adică distribuția nivelelor de gri la nivelul întregii imagini) este evident mixtura distribuțiilor individuale ale celor două tipuri de obiecte și se poate scrie ca:

$$h(i) = P_0 w_0(i) + P_1 w_1(i), \quad i = 0, 1, \dots, M-1. \quad (12.16)$$

Figura 12.6 prezintă o ilustrare grafică a erorii.

Pragul optim va minimiza eroarea de segmentare a pixelilor. Minimizarea erorii (12.15) conduce la rezolvarea ecuației în necunoscuta T :

$$\frac{\partial P_E(T)}{\partial T} = 0, \quad (12.17)$$

ceea ce conduce la relația:

$$-P_0 w_0(T) + P_1 w_1(T) = 0, \quad (12.18)$$

ce poate fi rescrisă ca:

$$\frac{w_1(T)}{w_0(T)} = \frac{P_0}{P_1}. \quad (12.19)$$

Forma (12.19) se poate compara cu forma finală a criteriului de decizie Bayes dată de (6.14). Constatăm astfel că ne aflăm într-un cadru de alegere simetrică a costurilor de decizie (ambele erori de decizie au același cost, ambele decizii corecte au același cost), astfel încât pragul testului Bayes este dat de raportul probabilităților a priori de apariție a celor două tipuri de obiecte.

Presupunerea că distribuția nivelelor de gri a diferitelor tipuri de obiecte este de tip normal (Gaussian) este relativ des întâlnită. În aceste condiții, distribuțiile $w_0(x)$ și $w_1(x)$

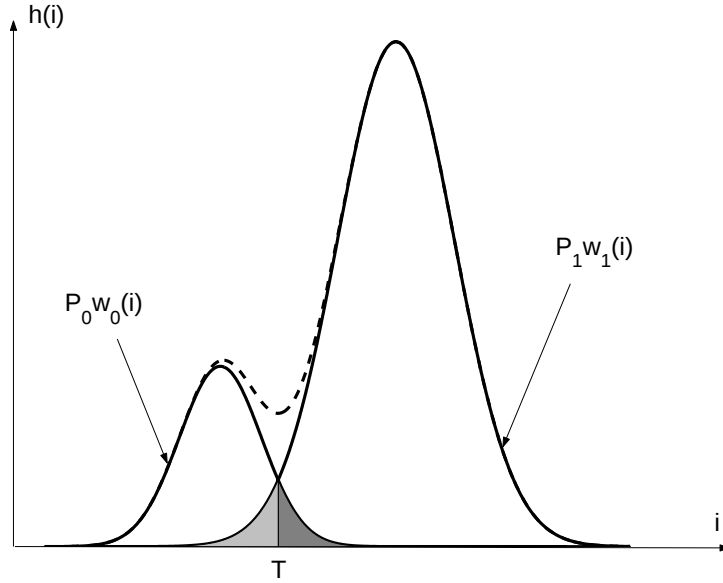


Figura 12.6: Eroarea de clasificare corespunzătoare separării a două moduri Gaussiene prin pragul T corespunde ariei hașurată din figură. Histograma imaginii, dată de (12.16), este figurată cu linie punctată.

sunt distribuții normale, $\mathcal{N}(m_0, \sigma_0)$ și $\mathcal{N}(m_1, \sigma_1)$, iar ecuația (12.19) devine:

$$\frac{\frac{1}{\sigma_1 \sqrt{2\pi}} \exp \left[-\frac{(T-m_1)^2}{2\sigma_1^2} \right]}{\frac{1}{\sigma_0 \sqrt{2\pi}} \exp \left[-\frac{(T-m_0)^2}{2\sigma_0^2} \right]} = \frac{P_0}{P_1} \quad (12.20)$$

Prin logaritmare, se obține o ecuație de gradul doi prin a cărei rezolvare se determină pragul necunoscut T :

$$T^2 \left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2} \right) - 2T \left(\frac{m_0}{\sigma_0^2} - \frac{m_1}{\sigma_1^2} \right) + \left(\frac{m_0^2}{\sigma_0^2} - \frac{m_1^2}{\sigma_1^2} \right) - 2 \ln \frac{P_0 \sigma_1}{P_1 \sigma_0} = 0. \quad (12.21)$$

Pentru cazul particular $\sigma_0 = \sigma_1 \stackrel{\text{not}}{=} \sigma$, ecuația (12.21) devine o ecuație de gradul unu, a cărei soluție este:

$$T = \frac{m_0 + m_1}{2} - \frac{\sigma^2}{m_0 - m_1} \ln \frac{P_0}{P_1} \quad (12.22)$$

Metoda se poate extinde și pentru imagini ce conțin mai mult de două tipuri de obiecte; în acest caz este însă necesară o presupunere suplimentară de localizare a modurilor, astfel încât să se poată considera că influența fiecărui mod este limitată la intervale nesuprapuse de nivele de gri și deci eroarea de clasificare a pixelilor imaginii se va referi doar la perechi de moduri vecine.

12.4 Compresia imaginilor cu transformate

Compresia imaginilor înseamnă reprezentarea acestora într-o formă simplificată, care necesită un spațiu mai mic de stocare sau o durată mai mică de transmisie. Dorim

reprezentarea aceleiași conținut vizual într-o formă mai compactă; noua reprezentare trebuie să permită refacerea identică (sau aproape identică) a conținutului imaginii inițiale – vom spune că avem de-a face, în cele două cazuri, cu o compresie fără pierderi (respectiv cu pierderi).

Posibilitatea de a stoca/transmite datele comprimat provine din redundanța existentă în succesiunea valorilor semnalului ce se comprimă/ compactează/ codează, redundanță ce are două cauze:

- Semnalele reale sunt corelate, ceea ce este echivalent cu faptul că valori apropiate ale semnalului depind una de alta. Astfel, prin codarea fiecărei valori independent de valorile alătuate se induce în semnal o redundanță care este cu atât mai mare cu cât corelația este mai puternică.
- Pentru majoritatea semnalelor, valorile acestora nu au o distribuție uniformă: anumite valori sunt mai probabile decât altele, fapt de care nu se ține în general cont, toate eșantioanele semnalului fiind codate cu același număr de biți, deși, după cum se știe din teoria codării, prin alocarea unui cod mai scurt valorilor mai probabile se poate reduce cantitatea globală de date necesare pentru stocarea întregului semnal.

Dintre tehnicile de bază de compresie bazate pe exploatarea corelației din semnale, cele mai comune sunt compresia cu transformate (discutată în capitolul 10) și cea predictivă (menționată în capitolul 9)⁴. Pentru eliminarea celei de-a doua cauze a redundanței, se folosește codarea entropică, de tip Huffman.

Eficiența unei metode de compresie se măsoară prin doi parametri: *raportul de compresie*, care este un indicator al reducerii cantității de informație necesare pentru stocarea imaginii, și *factorul de calitate*, care este o măsură a reducerii calității imaginii induse de compresie. Ca în multe alte probleme ingineresti, cele două cerințe sunt antagoniste (cu cât raportul de compresie este mai mare, cu atât factorul de calitate este mai scăzut, și vice-versa) proiectarea metodelor de compresie trebuind să asigure un compromis rezonabil între cele două.

Imaginile, în general, sunt semnale caracterizate de o corelație foarte puternică între valori vecine. Acest fapt se datorează existenței, în majoritatea imaginilor naturale, a unor zone uniforme extinse, zone caracterizate de valori similare ale pixelilor. Astfel, este de așteptat de la o metodă de compresie de imagini o reducere semnificativă a cantității de date, pentru factori de calitate mari.

Standardul de compresie JPEG, pe care-l vom prezenta schematic în continuare, face apel la toate tipurile de metode de compresie menționate mai sus (cu transformate, predictivă, entropică)

Compresia cu transformate a imaginilor se bazează pe proprietățile de conservare și compactare a energiei (și, implicit, de decorelare a componentelor) pe care le prezintă majoritatea transformărilor unitare discutate în capitolul 10. Atâta vreme cât cea mai

⁴Pentru imagini există multe alte tehnici specifice de compresie, dintre care cităm cuantizarea vectorială, compresia cu fractali etc.

mare parte a energiei este distribuită după transformarea valorilor originale ale imaginii în numai câteva componente transformate, toate celelalte valori transformate pot fi ignorate; astfel memoria necesară reprezentării este mult mai mică. Aplicarea practică a compresiei cu transformate are deci în vedere trei aspecte: alegerea transformatei bidimensionale, stabilirea numărului de valori ce se păstrează și, în fine, cuantizarea acestora.

După cum am discutat, există impedimente de ordin practic pentru utilizarea transformatei optimale Karhunen-Loève, drept pentru care în practică se folosesc aproximări ale acesteia. În condițiile în care majoritatea imaginilor naturale pot fi approximate printr-un model Markov de ordin unu puternic corelat (exprimând dependența puternică a valorii pixelilor de valorile vecinilor lor imediați), transformata cosinus discretă (prezentată pentru cazul unidimensional în paragraful 10.3) s-a dovedit cea mai bună alegere și a fost adoptată de standardul de compresie JPEG⁵. Transformarea cosinus discretă a unei imagini (deci bidimensională) se reduce la aplicarea succesivă a transformării unidimensionale pe fiecare dintre liniile imaginii și apoi pe fiecare coloană a matricii rezultate intermediar (pentru mai multe detalii vezi referința [6]).

Standardul de compresie JPEG⁶ (identificat de fișiere imagine cu extensia `.jpg` sau `.jpeg`) a fost definit în 1992 (ISO/ IEC 10928-1, ITU-T Rec. T-81). Imaginea este divizată în blocuri adiacente de 8×8 pixeli, iar fiecărui bloc i se aplică o transformată cosinus bidimensională⁷. Cei 64 de coeficienți ai transformării sunt aranjați într-un șir unidimensional prin baleierea elementelor matricii de 8×8 elemente într-o ordine prestabilită (un zig-zag pe diagonale paralele cu diagonala secundară) care face ca valorile cele mai semnificative cantitativ să ocupe primele poziții în vectorul rezultat. Coeficienții sunt apoi cuantizați cu pași de cuantizare diferiți în funcție de poziția lor în șir, pași prestabiliți prin standard. În mare, primele componente sunt cuantizate cu o precizie mai mare (deci cu un pas de cuantizare mai mic), în timp ultimele componente (cele mai puțin semnificative) sunt cuantizate grosier, cu cuante mari (ceea ce deseori duce la anularea completă a acestora). Calitatea compresiei este dată de calitatea cuantizării, întrucât pierderile totale introduse de compresia JPEG provin numai din cuantizarea valorilor transformate. Calitatea compresiei poate fi reglată de utilizator prin setarea unui factor de calitate Q care este un scalar cu valori între 1 și 100 cu care se scalează valorile pașilor de cuantizare standardizate. În mod evident pentru $Q = 1$ se obține calitatea cea mai bună (și compresia minimă), în timp ce pentru $Q \rightarrow 100$ se obține calitatea cea mai slabă (și compresia maximă). În figura 12.7 este prezentat rezultatul compresiei aceleiași imagini prin standardul JPEG pentru două valori ale factorului de calitate.

În continuare, valorile transformate și cuantizate ce corespund primei poziții din fiecare bloc de imagine (corespunzând valorii medii a nivelelor de gri din blocul respectiv) sunt

⁵În standardul de compresie JPEG2000, transformata cosinus discretă a fost înlocuită cu transformata wavelet, cu performanțe sporite

⁶Acronim pentru Joint Photographic Experts Group

⁷Calculul transformatei pe blocuri în loc de întreaga imagine a fost dictat de rațiuni ce țin de complexitatea algoritmului de codare și de obținerea unei statistici a valorilor cât mai apropiate de modelul markovian de ordinul unu.



Figura 12.7: Variante de compresie JPEG a imaginii cu nivele de gri din figura 12.4; imaginea din stânga este comprimată cu un factor de calitate mediu ($Q = 75$) la un raport de compresie de 35:1, iar imaginea din dreapta este comprimată cu un factor de calitate scăzut ($Q = 95$) ce corespunde unui raport de compresie de 48:1.

codate predictiv printr-o tehnică de tip DPCM (acronim pentru *Differential Pulse Code Modulation*⁸). În cele din urmă, șirul unidimensional de valori rezultat prin concatenarea vectorilor 64-dimensionali corespunzători fiecărui bloc de imagine de dimensiune 8×8 este codat entropic printr-o tehnică de codare Huffman.

Factorii de compresie ce rezultă sunt cuprinși în mod tipic între 10 și 100. Aspectul de compresie cu pierderi (diferențele față de imaginea originală) se manifestă prin efectul de *blocking*: sublinierea frontierelor de separație a blocurilor de bază (efect observabil și în figura 12.7).

⁸Modulația diferențială a impulsurilor în cod este o tehnică de codare predictivă simplificată a valorilor unui semnal: în locul fiecărui eșantion al semnalului este reținută/transmisă diferența între eșantionul respectiv și cel anterior, ceea ce este echivalentul unui predictor de ordinul unu, caracterizat de $\hat{\xi}(n+1) = \xi(n)$.

Anexa A

Impulsul Dirac

Să presupunem funcția $f(x)$ având expresia

$$f(x) = \begin{cases} E_1 & \text{dacă } x < x_1 \\ \frac{E_2 - E_1}{x_2 - x_1}(x - x_1) + E_1 & \text{dacă } x_1 \leq x \leq x_2 \\ E_2 & \text{dacă } x > x_2 \end{cases} . \quad (\text{A.1})$$

Graficul funcției f este prezentat în figura A.1.(a). Derivata funcției, prezentată în figura A.1.(b), este nulă pe intervalele pe care funcția e constantă, și este constantă pe intervalul pe care funcția crește liniar:

$$f'(x) = \begin{cases} 0 & \text{dacă } x < x_1 \text{ sau } x > x_2 \\ \frac{E_2 - E_1}{x_2 - x_1} & \text{dacă } x_1 \leq x \leq x_2 \end{cases} . \quad (\text{A.2})$$

Să presupunem în continuare că nivelele celor două paliere ale funcției $f(x)$, și anume E_1 și E_2 sunt constante. În schimb, să presupunem că facem să varieze limitele intervalului de creștere liniară a funcției, și anume x_1 și x_2 . Se observă că indiferent de cât de apropiate (sau îndepărtate) sunt x_1 și x_2 una de cealaltă, aria de sub graficul derivatei $f'(x)$ a funcției (aria hașurată în figura A.1.(b)) este constantă, de valoare $A = E_2 - E_1$.

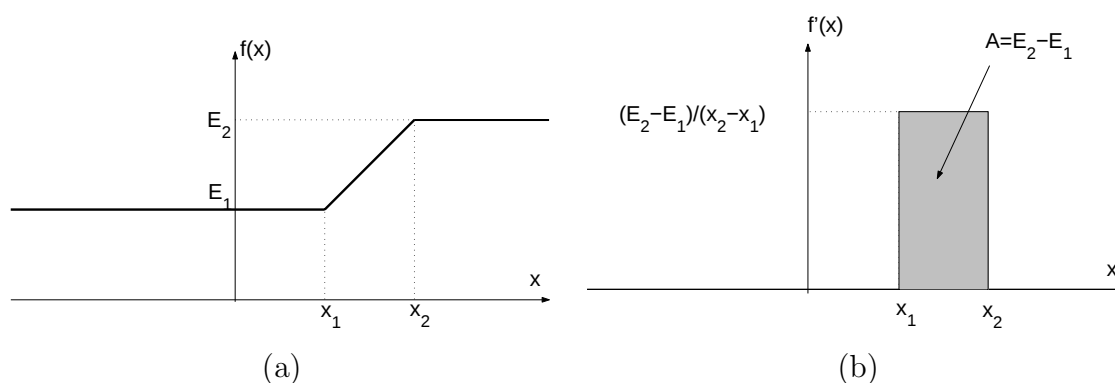


Figura A.1: Funcția $f(x)$ și derivata ei.

Considerăm acum forma aceleiași funcții și a derivatei sale în cazul limită $x_2 \rightarrow x_1$. În acest caz, funcția $f(x)$ efectuează un salt brusc de la nivelul E_1 la E_2 în punctul x_1 (v. figura A.2.(a)). Pe măsură ce x_2 se apropie de x_1 , intervalul $[x_1, x_2]$ se micșorează, dar valoarea derivatei pe intervalul respectiv crește. La limită, derivata $f'(x)$ capătă o valoare infinită, într-un singur punct, *dar continuă să aibă arie constantă*. Cu alte cuvinte, când $x_2 \rightarrow x_1$, în derivata funcției apare un impuls Dirac în punctul x_1 de arie $E_2 - E_1$ (v. figura A.2.(b)).

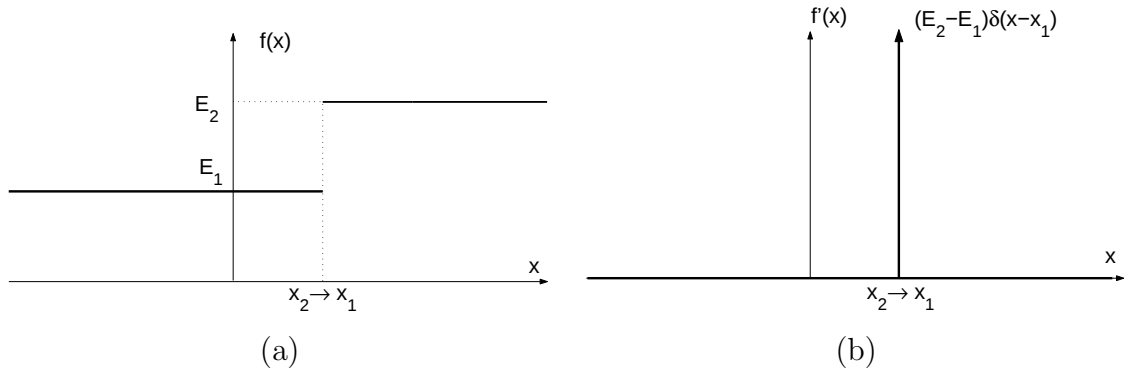


Figura A.2: Funcția $f(x)$ și derivata ei în cazul limită $x_2 \rightarrow x_1$.

Rezultă din cele de mai sus ca impulsul Dirac, notat cu $\delta(x)$, este un semnal având valoare infinită într-un singur punct¹, dar care are *arie finită* (aleasă prin convenție egală cu unitatea):

$$\delta(x) = \begin{cases} \infty & \text{dacă } x = 0 \\ 0 & \text{în rest} \end{cases}, \quad (\text{A.3})$$

cu

$$\int_{-\infty}^{\infty} \delta(x) dx = 1. \quad (\text{A.4})$$

Referitor la impulsul Dirac, se poate demonstra următoarea relație importantă:

$$\int_{-\infty}^{\infty} f(x) \delta(x - x_0) dx = \int_{x_0 - \epsilon}^{x_0 + \epsilon} f(x) \delta(x - x_0) dx \underset{\epsilon \searrow 0}{\approx} f(x_0) \underbrace{\int_{x_0 - \epsilon}^{x_0 + \epsilon} \delta(x - x_0) dx}_1 = f(x_0). \quad (\text{A.5})$$

Concluzia finală pe care o tragem în urma acestei expunerii este următoarea: dacă o funcție prezintă un salt (are o discontinuitate) într-un punct x_0 , atunci în derivata funcției, în punctul x_0 va apărea un impuls Dirac de arie egală numeric cu amplitudinea saltului efectuat de funcție.

¹Este evident că din punct de vedere strict matematic, o astfel de definiție a unui semnal care ia valori infinite este incorectă. Totuși, vom păstra definiția ca atare, întrucât ea ne permite să eludăm trecerea la limită când $x_2 \rightarrow x_1$ de fiecare dată.

Bibliografie

- [1] V. Buzuloiu. *Teoria transmisiunii informației*. Note de curs.
- [2] A. Spătaru. *Teoria transmisiunii informației*. Editura Didactică și Pedagogică, București, 1983.
- [3] A. Papoulis. *Probability, random variables and stochastic processes*. McGraw-Hill, New York, NY, 3rd edition, 1991.
- [4] S. Haykin. *Adaptive filter theory*. Prentice Hall, Englewood Cliffs, NJ, 2nd edition, 1991.
- [5] A. K. Jain. *Fundamentals of digital image processing*. Prentice Hall, Englewood Cliffs, NJ, 1989.
- [6] C. Vertan. *Prelucrarea și analiza imaginilor*. Editura Printech, București, 1999.
- [7] A. Spătaru. *Fondements de la théorie de la transmission de l'information*. Presses Polytechniques et Universitaires Romandes, Lausanne, 1987.
- [8] C. Vertan, I. Gavăt, și R. Stoian. *Variabile și procese aleatoare: principii și aplicații*. Editura Printech, București, 1999.
- [9] R. Boite și M. Kunt. *Traitement de la parole*. Presses Polytechniques Romandes, Lausanne, 1987.